0 44744

National Library of Canada
Canadian Theses Division
Ottawa, Canada
K1A 0N4

Bibliothèque nationale du Canada
Division des thèses canadiennes

PERMISSION TO MICROFILM — AUTORISATION DE MICROFILMER

- Please print or type — Écrire en lettres moulées ou dactylographier

Full Name of Author — Nom complet de l'auteur: PETER HUMPHREY HARASYM

Date of Birth — Date de naissance: NOV 16, 1941

Country of Birth — Lieu de naissance: CANADA

Permanent Address — Résidence fixe:
144 Rundlewood Close NE
Calgary, Alberta
T1Y 2P3

Title of Thesis — Titre de la thèse: The Analysis of Various Techniques Used For Scoring Patient Management Problems

University — Université: Alberta

Degree for which thesis was presented — Grade pour lequel cette thèse fut présentée: Ph. D.

Year this degree conferred — Année d'obtention de ce grade: 1980

Name of Supervisor — Nom du directeur de thèse: Dr. S. Hunka

Permission is hereby granted to the NATIONAL LIBRARY OF CANADA to microfilm this thesis and to lend or sell copies of the film.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

L'autorisation est, par la présente, accordée à la BIBLIOTHÈQUE NATIONALE DU CANADA de microfilmer cette thèse et de prêter ou de vendre des exemplaires du film.

L'auteur se réserve les autres droits de publication; ni la thèse ni de longs extraits de celle-ci ne doivent être imprimés ou autrement reproduits sans l'autorisation écrite de l'auteur.

Date: March 8, 1980

Signature: Peter Harasym

NL-91 (4/77)

National Library of Canada
Collections Development Branch

Canadian Theses on
Microfiche Service

Bibliothèque nationale du Canada
Direction du développement des collections

Service des thèses canadiennes
sur microfiche

NOTICE

The quality of this microfiche is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us a poor photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this film is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30. Please read the authorization forms which accompany this thesis.

THIS DISSERTATION
HAS BEEN MICROFILMED
EXACTLY AS RECEIVED

AVIS

La qualité de cette microfiche dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de mauvaise qualité.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, examens publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de ce microfilm est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30. Veuillez prendre connaissance des formules d'autorisation qui accompagnent cette thèse.

LA THÈSE A ÉTÉ
MICROFILMÉE TELLE QUE
NOUS L'AVONS REÇUE

Ottawa, Canada
K1A 0N4

NL-339 (Rev. 8/80)

THE UNIVERSITY OF ALBERTA

THE ANALYSIS OF VARIOUS TECHNIQUES USED FOR SCORING PATIENT MANAGEMENT PROBLEMS

by

© PETER H. HARASYM

A THESIS
SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE
OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF EDUCATIONAL PSYCHOLOGY

Edmonton, Alberta

Spring, 1980

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, A Thesis entitled "THE ANALYSIS OF VARIOUS TECHNIQUES USED FOR SCORING PATIENT MANAGEMENT PROBLEMS" submitted by PETER H. HARASYM in partial fulfilment of the requirements for the degree of Doctor of Philosophy.

S. Hunka
Supervisor

C. B. Hazlett

E. W. Romaniuk

Date November 5, 1979

Gordon Page
External Examiner

ABSTRACT

An investigation was undertaken to determine whether different scoring procedures have varying effects upon examinee computer patient management problem (CPMP) scores. One hundred and eleven, fourth year medical students were examined on four CPMPs. Student responses were scored by twelve scoring procedures (keys), four of which are used extensively by licensing agents and/or medical schools. The analysis of the data indicated that the weightings for the same options could vary greatly over scoring keys (*i.e.*, from indispensible positive to unforgiveable negative). Scoring keys also varied in the number and proportion of marks allocated to positive and negative options. These variations resulted in alterations to the:

1) shape of the distribution of scores,
2) score variance,
3) trait or behavior being measured,
4) mean scores,
5) examinee satisfactory/unsatisfactory status,
6) test/retest reliability, and
7) rank ordering of examinees.

Based upon insights and results from this investigation, it was recommended that a profile of examinee clinical performance be generated, that the expert problem-solvers' performance be used for determining and/or validating option weightings, that differential weights be used to reflect the importance of an options's contribution to resolving the patient's problem

and that continued efforts be directed towards furthering our understanding and skills in measuring the complex and elusive process of clinical problem-solving.

ACKNOWLEDGEMENTS

The author wishes to express his sincere thanks to the members of his committee, Dr. C. B. Hazlett, Dr. S. Hunka, Dr. G. G. Page, Dr. M. W. Petruk and Dr. E. W. Romaniuk. Special thanks are due to Dr. S. Hunka, the committee chairman, for his patience and insightful leadership.

Appreciation is expressed to the R. S. McLaughlin Examination and Research Center and the many physicians and medical students who helped make this thesis possible.

Special thanks are also due to Ms. Heather Staple for her constant encouragement, patience and understanding and for having spent many unselfish hours typing this manuscript. It is to her and my daughter, Lisa Harasym, that this thesis is dedicated.

This project was funded by a Fellowship from the International Business Machine Corporation of Canada. The ideas or views expressed in this manuscript may not necessarily be those endorsed by the funding agent.

TABLE OF CONTENTS

CHAPTER Page

I. PURPOSE OF STUDY 1

  Importance of Study 3

  Scope of Study 4

II. RELATED LITERATURE AND RESEARCH 7

  Medical Inquiry 7

  Clinical Competence 14

  Simulations, Techniques Used to Assess Clinical Competence 16

  Scoring Procedures 25

III. CURRENT SCORING PROCEDURES 29

  Scoring Procedures Currently Used 31

    Group consensus, constant weight, single key, and T/F method 31

    Group consensus, differential weights, single key, and sum method 32

    Group performance, constant weight, single key, and sum method 34

    Group consensus and performance, differential weights, single key, and sum method 36

  Comparison of Current Scoring Procedures . . 38

    Linear versus branching designs 38

    Expert consensus and/or performance . . . 41

    Constant or differential weights 46

    True/False or sum method 48

CHAPTER Page

Single or multiple keys 55

Scoring formulas 57

Summary 59

IV. SCORING PROCEDURES 60

Classification 60

Notation 61

Clinical decisions in a computer simulation of a patient management problem 61

Categorization of clincial decisions . . . 61

Weights of clinical decisions. 62

Options selected 62

Eleven Scoring Procedures 63

Group consensus, constant weights, and single key (GCS) 63

Group consensus, differential weights, and single key (GDS) 65

Individual judgements, constant weights, and single key (ICS) 67

Individual judgements, differential weights, single key (IDS) 67

Individual judgements, constant weights, multiple key (ICM) 68

Individual judgement, differential wieghts, multiple key (IDM) 69

Computer performance, constant weights, single key (CCS) 69

Computer performance, differential weights,, single key (CDS) 71

Computer performance, constant weights, multiple key (CCM) 71

CHAPTER Page

Computer performance, differential weights, multiple key (CDM) 72

Group consensus and computer performance, differential weights and single key (Mc) 72

Author, differential weights, single key (author) 74

Assumptions and Limitations Underlying Scoring Procedures 76

Method of categorization 76

Group consensus (i.e., GCS and GDS) . . 76

Individual judgements (i.e., ICS, IDS, ICM and IDM) 77

Computer performance (i.e., CCS, CDS, CCM and CDM) 77

Weights 78

Constant (i.e., GCS, ICS, ICM CCS and CCM) 78

Differential (i.e., Author, GDS, IDS, IDM, CDS, CDM and Mc). 78

Key 79

Single (i.e., Author, GCS, GDS, ICS, IDS, CCS, CDS and Mc) 79

Multiple (i.e., ICM, IDM,CCM and CDM) 79

Summary 80

V. MATERIALS AND METHODS 81

Subjects 81

Examinations (CPMPs) 81

CPMP 1 82

CPMP 2 84

CHAPTER Page

CPMP 3 86

CPMP 4 89

Comparison of CPMPs 89

"Expert" Physicians 92

The Development of Scoring Keys 93

Description of Data 96

Method of Data Analysis 97

Reliability measures 97

Consistency across occasion 99

Estimation of a minimal standard of performance (MPL) 101

Estimating the reliability of homogeneous options with CPMPs 102

Validity measures 103

VI. RESULTS 104

Characteristics of Scoring Keys 104

Descriptive Statistics of the Examinee CPMP Scores 108

Inference about the variance among proficiency scores calculated using the twelve scoring procedures. . . 109

Skewness and kurtosis 126

Factor Analysis 127

Component structure underlying proficiency scores 127

Component analysis of proficiency scores on each CPMP 131

CPMP 1 131

CPMP 2 133

CHAPTER Page

CPMP 4 135

Discussion of the component analytic investigation of proficiency scores. . . 139

Component structure underlying error of commission scores 141

Component analysis of error of commission scores on each CPMP 144

CPMP 1 144

CPMP 2 146

CPMP 3 148

CPMP 4 150

Discussion of the factor analytic investigation of error of commission scores 152

Component structure underlying error of omission scores 154

Component analysis of error of omission scores on each CPMP 157

CPMP 1 158

CPMP 2 159

CPMP 3 162

CPMP 4 164

Discussion of component analytic investigation of error of omission scores 166

Component structure underlying efficiency scores 168

Component analysis of efficiency scores on each CPMP 171

CPMP 1 172

CPMP 2 174

CHAPTER Page

CPMP 3 176

CPMP 4 178

Discussion of the component analytic investigation of efficiency scores . . . 180

Summary of the component structure underlying CPMP scores 181

Multivariate Analysis 186

Multivariate analysis of CPMP scores . . . 186

Statistical analysis 187

Calculation of the generalized inverse of a large design matrix 187

One-way multivariate analysis 188

Linear model 188

Hypothesis 189

Results 190

Simultaneous-paired comparisons 191

CPMP 1, proficiency scores 194

CPMP 1, error of omission scores . . . 197

CPMP 1, efficiency scores 200

CPMP 2, proficiency scores 203

CPMP 2, error of omission scores . . . 206

CPMP 2, efficiency scores 209

CPMP 3, proficiency scores 212

CPMP 3, error of omission scores . . . 212

CPMP 3, efficiency scores 217

CPMP 4, proficiency scores 220

CPMP 4, error of omission scores . . . 223

CPMP 4, efficiency scores 226

CHAPTER Page

Summary and discussion of results obtained in the one-way multivariate analysis 229

Effects due to method of categorization 229

Effect of constant and differential weights upon examinee CPMP mean scores 238

Effect of single and multiple keys upon examinee CPMP mean scores 241

Number of Examinees Who Achieved Satisfactory Status on CPMPs 1-4 241

Rank Ordering of Examinee Scores 244

Reliability of Expert Judgements 250

Estimates of Reliability for Homogeneous Options Within CPMPs 251

Reliability estimates for CPMP 1 256

Reliability estimates for CPMP 2 256

Reliability estimates for CPMP 3 256

Reliability estimates for CPMP 4 257

Validity Measure 258

VII. SUMMARY, IMPLICATIONS, LIMITATIONS AND RECOMMENDATIONS 260

Summary 260

Shape of the distribution of scores . . . 261

Score variance 261

Trait or behavior measured 262

Mean scores 263

Examinee satisfactory (pass)/ unsatisfactory (fail) status 264

CHAPTER Page

Test/retest reliability 264

Rank ordering of examinee scores . . . 265

Conclusions 265

Implications 267

Limitations 273

Recommendations 273

REFERENCES 275

APPENDICES

A Sample of Linear Simulated Patient Encounter 279

B Meningitis Management Section 280

C Scoring Formulas for Simulation Exercises 282

D Instructions for Categorizing Options for Linear Problems 283

E Instructions for Categorizing Options for Branching Problems 284

F Distribution of Scores 285

G Algebraic Solution for Generalized Inverse of X'X matrix 297

H Sums of Squares and Cross Products due to Total (Y'Y matrix) 300

I Sums of Squares and Cross Products due to Model ($\hat{Y}'\hat{Y}$ matrix) 301

J Sums of Squares Due to Model Corrected for Sums of Squares Due to Means 302

K Sums of Squares and Cross-Products Due to Scoring Procedures (H matrix) 303

Page

L Effects Model (Beta weights due to scoring procedures) (B matrix) 304

M Sums of Squares and Cross-Products Due to Error (E matrix) 305

LIST OF TABLES

Table	Description	Page
3.1	Example of Scoring Key Based on Group Performance Constant Weight, Single Key, and Sum Method	35
3.2	Categorization of Proposed Scoring Procedures	39
4.1	Eleven Scoring Procedures Created by Varying 3 Methods of Categorizing Decisions, 2 Methods of Weighting and 2 Types of Keys	64
4.2	Categorization of Proposed Scoring Procedures	75
5.1	Biographical Data of "Expert " Physicians . .	93
5.2	Tasks and CPMPs Performed by the Three Groups of "Expert" Physicians	95
6.1	Number of Negative/Positive Options by CPMP and Scoring Procedure	106
6.2	Percentage Weight Out of a Scale of 100 Percent Related to Negative Options . . .	107
6.3	Relationship Between Structure of CPMP and Percentage Weight Allocated to Negative Options	108
6.4	Descriptive Statistics of Distribution of Proficiency Scores Calculated on CPMP 1 Using 12 Scoring Procedures	110
6.5	Descriptive Statistics of Distribution of Error of Omission Scores Calculated on CPMP 1 Using 12 Scoring Procedures	111
6.6	Descriptive Statistics of Distribution of Efficiency Scores Calculated on CPMP 1 Using 12 Scoring Procedures	112

Table	Description	Page
6.7	Descriptive Statistics of Distribution of Proficiency Scores Calculated on CPMP 2 Using 12 Scoring Procedures	113
6.8	Descriptive Statistics of Distribution of Error of Omission Scores Calculated on CPMP 2 Using 12 Scoring Procedures . . .	114
6.9	Descriptive Statistics of Distribution of Efficiency Scores Calculated on CPMP 2 Using 12 Scoring Procedures	115
6.10	Descriptive Statistics of Distribution of Proficiency Scores Calculated on CPMP 3 Using 12 Scoring Procedures	116
6.11	Descriptive Statistics of Error of Commission Scores Calculated on CPMP 3 Using 12 Scoring Procedures	117
6.12	Descriptive Statistics of Distribution of Efficiency Scores Calculated on CPMP 3 Using 12 Scoring Procedures	118
6.13	Descriptive Statistics of Distribution of Proficiency Scores Calculated on CPMP 4 Using 12 Scoring Procecures	119
6.14	Descriptive Statistics of Distribution of Error of Commission Scores Calculated on CPMP 4 Using 12 Scoring Procedures . . .	120
6.15	Descriptive Statistics of Distribution of Efficiency Scores Calculated on CPMP 4 Using 12 Scoring Procedures	121
6.16	T Test of Comparison of Variance Among Proficiency Scores Calculated by the Twelve Scoring Procedures on CPMP 1	122
6.17	T Test Comparison of Variance Among Proficiency Scores Calculated by the Twelve Scoring Procedures on CPMP 2	123
6.18	T Test Comparison of Variance Among Proficiency Scores Calculated by the Twelve Scoring Procedures on CPMP 4	124
6.19	T Test [illegible]rison of Variance Among Profic[illegible]cy Scores Calculated by the Twelve Scoring Procedures on CPMP 3	125

Table	Description	Page
6.20	Correlation Coefficient Between the Proficiency Scores of the 4 CPMPs Calculated Using the 12 Scoring Procedures	128
6.21	Varimax Rotated Principal Component Factor Analysis of Proficiency Scores . . .	129
6.22	Correlation Coefficient Between Proficiency Scores on CPMP 1 Calculated Using the 12 Scoring Procedures (N = 111)	132
6.23	Principal Component Factor Analysis of Proficiency Scores on CPMP 1	132
6.24	Correlation Coefficient Between Proficiency Scores on CPMP 2 Calculated Using the 12 Scoring Procedures (N = 111)	133
6.25	Varimax Rotated Principal Component Factor Analysis of Proficiency Scores on CPMP 2	134
6.26	Correlation Coefficient Between Proficiency Scores on CPMP 3 Calculated Using the 12 Scoring Procedures (N = 111)	135
6.27	Varimax Rotated Principal Component Factor Analysis of Proficiency Scores on CPMP 3	136
6.28	Correlation Coefficient Between Proficiency Scores on CPMP 4 Calculated Using the 12 Scoring Procedures (N = 111)	137
6.29	Varimax Rotated Principal Component Factor Analysis of Proficiency Scores on CPMP 4	138
6.30	Correlation Coefficient Between Errors of Commission Calculated Using 12 Scoring Procedures and 4 CPMPs	142
6.31	Varimax Rotated Principal Component Factor Analysis of Error of Commission Scores. . .	143

Table	Description	Page
6.32	Correlation Coefficient Between Errors of Commission on CPMP 1 Calculated Using the 12 Scoring Procedures (N = 111)	145
6.33	Principal Component Factor Analysis of Errors of Commission Scores on CPMP 1 . . .	145
6.34	Correlation Coefficient Between Errors of Commission on CPMP 2 Calculated Using the 12 Scoring Procedures (N = 111)	146
6.35	Varimax Rotated Principal Component Factor Analysis of Errors of Commission Scores on CPMP 2	147
6.36	Correlation Coefficient Between Errors of Commission on CPMP 3 Calculated Using the 12 Scoring Procedures (N = 111)	148
6.37	Varimax Rotated Principal Component Factor Analysis of Errors of Commission Scores on CPMP 3	149
6.38	Correlation Coefficient Between Errors of Commission on CPMP 4 Calculated Using the 12 Scoring Procedures (N = 111)	151
6.39	Varimax Rotated Principal Component Factor Analysis of Error of Commission Scores on CPMP 4	151
6.40	Correlation Coefficient Between Errors of Omission Calculated Using 12 Scoring Procedures and 4 CPMPs	155
6.41	Varimax Rotated Principal Component Factor Analysis of Error of Omission Scores . . .	156
6.42	Correlation Coefficient Between Errors of Omission on CPMP 1 Calculated Using the 12 Scoring Procedures (N = 111)	158
6.43	Principal Component Factor Analysis of Error of Omission Scores on CPMP 1	159
6.44	Correlation Coefficient Between Errors of Omission on CPMP 2 Calculated Using the 12 Scoring Procedures (N = 111)	160

Table Description Page

6.45 Varimax Rotated Principal Component Factor Analysis of Error of Omission Scores on CPMP 2 161

6.46 Correlation Coefficient Between Errors of Omission on CPMP 3 Calculated Using the 12 Scoring Procedures (N = 111) . . . 162

6.47 Varimax Rotated Principal Component Factor Analysis of Error of Omission Scores on CPMP 3 163

6.48 Correlation Coefficient Between Errors of Omission on CPMP 4 Calculated Using the 12 Scoring Procedures (N = 111) . . . 164

6.49 Varimax Rotated Principal Component Factor Analysis of Error of Omission Scores on CPMP 4 165

6.50 Correlation Coefficient Between Efficiency Scores Calculated Using 12 Scoring Procedures on 4 CPMPs 169

6.51 Varimax Rotated Principal Component Factor Analysis of Efficiency Scores 170

6.52 Correlation Coefficient Between Efficiency Scores on CPMP 1 Calculated Using the 12 Scoring Procedures (N = 111) 172

6.53 Varimax Rotated Principal Component Factor Analysis of Efficiency Scores on CPMP 1 173

6.54 Correlation Coefficient Between Efficiency Scores on CPMP 2 Calculated Using the 12 Scoring Procedures (N = 111) 174

6.55 Varimax Rotated Principal Component Factor Analysis of Efficiency Scores on CPMP 2 175

6.56 Correlation Coefficient Between Efficiency Scores on CPMP 3 Calculated Using the 12 Scoring Procedures (N = 111) 176

6.57 Varimax Rotated Principal Component Factor Analysis of Efficiency Scores on CPMP 3 177

Table	Description	Page
6.58	Correlation Coefficient Between Efficiency Scores on CPMP 4 Calculated Using the 12 Scoring Procedures (N = 111)	178
6.59	Varimax Rotated Principal Component Factor Analysis of Efficiency Scores on CPMP 4 . .	179
6.60	Number of Factors and Structure of CPMP Scores	184
6.61	Roas Approximate F Test Using Wilks Lambda - Ray's Maximum Eigenvalue Test . .	191
6.62	Multiple Comparison of Mean Proficiency Scores on CPMP 1 Calculated Using the 12 Scoring Procedures	195
6.63	Multiple Comparison of Mean Proficiency Scores on CPMP 1 By: A. Method of Categorization B. Constant Versus Differential Weights C. Single Versus Multiple Keys	196
6.64	Multiple Comparison of Mean Errors of Omission Scores on CPMP 1 Calculated Using the 12 Scoring Procedures	198
6.65	Multiple Comparison of Mean Error of Omission Scores on CPMP 1 By: A. Method of Categorization B. Constant Versus Differential Weights C. Single Versus Multiple Key	199
6.66	Multiple Comparison of Mean Efficiency Scores on CPMP 1 Calculated Using the 12 Scoring Procedures	201
6.67	Multiple Comparison of Mean Efficiency Scores on CPMP 1 by: A. Method of Categorization B. Constant Versus Differential Weights C. Single Versus Multiple Key	202
6.68	Multiple Comparison of Mean Proficiency Scores on CPMP 2 Calculated Using the 12 Scoring Procedures	204
6.69	Multiple Comparison of Mean Proviciency Scores on CPMP 2 by: A. Method of Categorization B. Constant Versus Differential Weights C. Single Versus Multiple Key	205

Table	Description	Page
6.70	Multiple Comparison of Mean Errors of Commission Scores on CPMP 2 Calculated Using the 12 Scoring Procedures	207
6.71	Multiple Comparison of Mean Error of Omission Scores on CPMP 2 by: A. Method of Categorization B. Constant Versus Differential Weights C. Single Versus Multiple Keys	208
6.72	Multiple Comparison of Mean Efficiency Scores on CPMP 2 Calculated Using the 12 Scoring Procedures	210
6.73	Multiple Comparison of Mean Efficiency Scores on CPMP 2 by: A. Method of Categorization B. Constant Versus Differential Weights C. Single Versus Multiple Keys	211
6.74	Multiple Comparison of Mean Proficiency Scores on CPMP 3 Calculated Using the 12 Scoring Procedures	213
6.75	Multiple Comparison of Mean Proficiency Scores on CPMP 3 by: A. Method of Categorization B. Constant Versus Differential Weights C. Single Versus Multiple Keys	214
6.76	Multiple Comparison of Mean Errors of Commission Scores on CPMP 3 Calculated Using the 12 Scoring Procedures	215
6.77	Multiple Comparison of Mean Erorr of Omission Scores on CPMP 3 by: A. Method of Categorization B. Constant Versus Differential Weights C. Single Versus Multiple Keys	216
6.78	Multiple Comparison of Mean Efficiency Scores on CPMP 3 Calculated Using the 12 Scoring Procedures	218
6.79	Multiple Comparison of Mean Efficiency Scores on CPMP 3 by: A. Method of Categorization B. Constant Versus Differential Keys C. Single Versus Multiple Keys	219

Table Description Page

6.80 Multiple Comparison of Mean Proficiency Scores on CPMP 4 Calculated Using the 12 Scoring Procedures 221

6.81 Multiple Comparison of Mean Proficiency Scores on CPMP 4 by: A. Method of Categorization B. Constant Versus Differential Weights C. Single Versus Multiple Keys 222

6.82 Multiple Comparison of Mean Error of Commission Scores on CPMP 4 Calculated Using the 12 Scoring Procedures 224

6.83 Multiple Comparison of Mean Error of Omission Scores on CPMP 4 by: A. Method of Categorization B. Constant Versus Differential Weights C. Single Versus Multiple Keys 225

6.84 Multiple Comparison of Mean Efficiency Scores on CPMP 4 Calculated Using the 12 Scoring Procedures 227

6.85 Multiple Comparison of Mean Efficiency Scores on CPMP 4 by: A. Method of Categorization B. Constant Versus Differential Weights C. Single Versus Multiple Keys 228

6.86 Summary of Multiple Comparison of Mean Proficiency Scores on CPMP 1-4 Calculated Using Scoring Procedures with Differential Weights and Single Keys 230

6.87 Summary of Multiple Comparison of Mean Proficiency Scores on CPMPs 1-4 Calculated Using Scoring Procedures With Constant Weights and Single Keys 231

6.88 Summary of Multiple Comparison of Mean Error of Omission Scores on CPMPs 1-4 Calculated Using Scoring Procedures With Differential Weights and Single Keys . . . 231

6.89 Summary of Multiple Comparison of Mean Error of Omission Scores on CPMP 1-4 Calculated Using Scoring Procedures With Constant Weights and Single Keys 231

Table	Description	Page
6.90	Summary of Multiple Comparison of Mean Efficiency Scores on CPMPs 1-4 Calculated Using Scoring Procedures With Differential Weights and Single Keys	232
6.91	Summary of Multiple Comparison of Mean Efficiency Scores on CPMPs 1-4 Calculated Using Scoring Procedures With Constant Weights and Single Keys	232
6.92	Number of Significant Paired Comparisons of Mean Proficiency, Error of Omission, and Efficiency Scores on CPMPs 1-4 and the Corresponding 95% Confidence Interval	234
6.93	Number of Positive Options Identified Within Eight Scoring Procedures and Average Number of Selections by Examinee and Expert Problem Solver . .	236
6.94	Correlation Coefficient Between Scoring Key and Frequency that Options were Selected in CPMPs 1-4 by Expert Problem-Solver and Examinees	236
6.95	Summary of Multiple Comparison of Mean Scores on CPMPs 1-4 Calculated Using Scoring Procedures With Constant and Differential Weights	239
6.96	Summary of Multiple Comparison of Mean Scores on CPMPs 1-4 Calculated Using Scoring Procedures With Single and Multiple Keys	242
6.97	Number of Examinees (N = 111) Who Achieved Satisfactory Status on CPMPs 1-4 Scored by 12 Different Scoring Procedures	243
6.98	Number of Disagreements in Satisfactory (Pass)/ Unsatisfactory (Fail) Status on CPMP 2 by Scoring Procedure	245
6.99	Rank Ordering (Spearman's Rho) of Students (N = 111) by Proficiency Scores Calculated on CPMP 1 Using 12 Scoring Procedures	246

Table	Description	Page
6.100	Rank Ordering (Spearman's Rho) of Students (N = 111) by Proficiency Scores Calculated on CPMP 2 Using 12 Scoring Procedures	247
6.101	Rank Ordering (Spearman's Rho) of Students (N = 111) By Proficiency Scores Calculated on CPMP 3 Using 12 Scoring Procedures	248
6.102	Rank Ordering (Spearman"s Rho) of Students (N = 111) by Proficiency Scores Calculated on CPMP 4 Using 12 Scoring Procedures	249
6.103	Reliabilities of Individual Judgements on Computer Selections on CPMPs 1-4	251
6.104	Reliability Estimates for Proficiency, Error of Commission and Error of Omission Scores Calculated Using Twelve Scoring Procedures on CPMP 1	252
6.105	Reliabliity Estimates for Proficiency, Error of Commission and Error of Omission Scores Calculated Using Twelve Scoring Procedures on CPMP 2	253
6.106	Reliability Estimates for Proficiency, Error of Commission and Error of Omission Scores Calculated Using Twelve Scoring Procedures on CPMP 3	254
6.107	Reliability Estimates for Proficiency, Error of Commission and Error of Omission Scores Calculated Using Twelve Scoring Procedures on CPMP 4	255
6.108	Number of Expert Problem-Solvers With Scores Above the Minimal Pass Level (MPL)	259

LIST OF FIGURES

Figure	Description	Page
1.1	Linear PMP	20
1.2	Diagram of a Branching PMP	22
3.1	Difference in Slope Between (+) and (-) Weights Assigned in Meningitis Problem	45
3.2	The Total Scores Using the National Board or the T/F Method Produces a Score of 9	49
3.3	Comparison of T/F and Sum Scores	50
3.4	Linear Relationship Between Scores Calculated Using True/False and Sum Methods	52
3.5	Linear Relationship Between % Scores Calculated Using Sum and True/False Method	53
5.1	Flowchart of CPMP 1 - A 44 Year Old Man with a Cardiac Problem	83
5.2	Flowchart of CPMP 2 - A 56 Year Old Man with Anemia of Unknown Origin	85
5.3	Flowchart of CPMP 3 - A 25 Year Old Female with a Gynecological Problem	87
5.4	Flowchart of CPMP 4 - A 24 Year Old Female with an Obstetrical Problem	90

CHAPTER I

PURPOSE OF STUDY

Accurate evaluation of professional skills is necessary to society to ensure a high standard of professional services. As a prerequisite to high standards of professional services, society must have institutions of higher learning which teach the necessary knowledge and evaluate the skills acquired by students. Licensing agencies are needed that can examine professionals, either at the time the professional first claims qualification, and/or throughout the duration of professional practice. It is the responsibility of the educational institutions and licensing agencies to develop and administer examinations of the highest quality which accurately assess predefined skills and knowledge. In order to develop high quality examinations, institutions and licensing agencies must clearly define what is being tested and constantly re-examine and improve their assessment instruments. In addition, they must explore, develop and refine new assessment methods.

Prior to 1950, the majority of assessment instruments contained multiple-choice questions that tended to measure the recall of factual information. In an effort to improve communication among educators and to help teach, classify and assess higher educational objectives, a taxonomy classifying cognitive skills was developed and described in a book edited by Benjamin S. Bloom (1956). Educators, in response to Bloom's taxonomy, became more aware of the complex thinking skills and of the inadequacies of their examination methods. Other evaluation techniques were sought to measure higher-level thinking skills and were found through developments in computer and simulation technology.

The medical profession was among the first to recognize and explore the potentials of patient simulations for teaching and assessing complex clinical decision making skills. Pencil and paper patient simulation were first developed and used successfully by medical schools (McGuire and Babbott, 1967). The advantages of using patient simulations over oral examinations were quickly recognized by licensing bodies and concerted efforts were made to develop patient management problems (Hubbard, 1971). With the advent of computers, the medical profession took advantage of such capabilities to simulate complex patient-physician encounters which were impossible using the earlier pencil and paper techniques. A new dimension in patient simulation was launched which today is generally referred to as computer patient management problem (CPMP).

However, with the introduction of clinical simulations many unforseen problems were introduced. One problem that has eluded investigation is that of scoring. This study concerned itself with the collection and scoring of clinical performance data using CPMPs. The principal focus is one of measurement to further the understanding of the effects that various scoring procedures have upon measures for evaluating clinical decisions.

1. Importance of Study

A large number of medical schools are using patient simulations to teach and evaluate students' clinical problem-solving skills. It is important that the scoring procedures accurately reflect students' capabilities since many educational, administrative and career decisions are made on the basis of the examinee scores. For example, the examinee scores could be used to identify learner strengths and weaknesses and used to guide the students' learning activities. The class scores when analyzed can be used by the teacher to revise the curriculum, learning experiences, or the testing instrument.

Licensing agencies are using patient simulations to certify candidates. Since the score generated determines the candidates that are licensed, it is important to the candidates, the licensing agency, and society that the scoring procedure accurately reflects the candidate's capabilities.

Researchers are using patient simulations to study and understand the cognitive processes associated in clinical problem-solving. In order to advance this body of knowledge it is necessary that scoring procedures accurately reflect subjects' activities. Incorrect scores may lead to incorrect research findings and conclusions.

Over the last twenty years many types of patient simulations and scoring procedures have been developed by medical schools, licensing agencies and researchers. Unfortunately there seems to be some confusion as to which scoring procedure should be selected. Different scoring procedures are used with the same type of patient simulations, and vise-versa. The mixing of scoring procedures exists in spite of the possibility that different scoring procedures could induce differences in examinee scores. The investigation of the effect of scoring procedures is overdue, considering the number of important decisions that are made on the examinee scores.

2. Scope of the Study

Edwards and Cronbach (1952) distinquish between two types of research: (1) survey research and (2) critical research. Survey research is undertaken when the investigator is relatively uncertain of the possible relationship among variables. The aim of the research is to determine the relationship between variables. On the other hand, critical research is undertaken when theoretical considerations indicate the questions to be asked and even indicate expected answers. The aim of theoretical research is to substantiate and, if necessary, alter the theoretical model or conceptions through observed data. Using Edwards' and Cronbach's definition the present study may be classified as survey research.

In order to meaningfully examine the effects various scoring procedures have upon evaluating clinical decision-making skills, it was necessary to carry out the following steps:

(1) review related medical literature on how physicians conduct a medical work-up from the initial patient encounter to the reaching of a final clinical decision,

(2) review the types of patient simulations used to investigate and evaluate clinical decision-making,

(3) define clinical competence,

(4) review scoring procedures used to quantify the appropriateness of clinical decisions made on patient simulated encounters,

7

(5) develop a classification system for categorizing procedures,

(6) develop new scoring procedures by varying the categories of the scoring classification system,

(7) devise scoring keys using data gathered from expert physicians,

(8) gather data of examinee clinical decisions using computer presented patient management problems,

(9) calculate examinee scores using the various

scoring keys,

(10) analyze examinee scores to determine the source of variation introduced by each component of the scoring procedure,

(11) establish critical scores which reflect clinical competence and incompetence,

(12) determine the extent to which various scoring procedures will affect candidates competence and incompetence status, and

(13) assess the internal validity of the scoring procedures by analysis of experts clinical decisions.

The above steps were followed to determine the effect various scoring procedures have upon examinee scores.

The following chapter reviews the research which describes how physicians conduct medical work-ups from the initial patient encounter to the final diagnosis and treatment. In addition, the various patient simulations developed to investigate and evaluate clinical problem solving are reviewed.

CHAPTER II

RELATED LITERATURE AND RESEARCH

1. Medical Inquiry

Central to the effective delivery of health care by the physician is the complex skill of clinical problem solving. The accuracy of this skill is crucial to the life and well-being of the patient. Since the primary objective of the simulated patient-management problem is to model the physician-patient encounter and to assess the accuracy of the physician's clinical problem-solving skills, it is first necessary to understand how physicians conduct a medical work-up from the initial encounter to the reaching of a final clinical judgement.

Recent studies and theories in the area of clinical judgement may be divided into two general types. The first relies on introspection to elucidate the mental processes by which the clinician solves problems. This procedure is exemplified by the work of Kleinmuntz (1968), Simon (1971), Barrows and Bennett (1972), Elstein (1972), and Shulman (1974). The second type also uses introspection, but rather than elucidating the mental processes, statistical models are used to replicate the judgement of the clinician without necessarily reproducing the cognitive steps. This approach is exemplified by the work of Hoffman (1960), Hammond *et*. *al*., (1964), and Goldberg (1970).

The most common method used to investigate the diagnostic process under controlled but natural conditions is to have actors and actresses play the role of patients. Laboratory data is supplied on lab report slips or real x-rays, and physical examinations are performed whenever possible on simulated patients. Insight into the physicians thinking is obtained by analyzing data gathered by (1) videotaping the physicians during the clinical encounter, (2) having physicians think aloud and (3) having physicians view their videotape in order to stimulate detailed recall of their intellectual cognitive processes. This method, used in whole or modified form produced surprisingly similar findings in spite of having studied physicians from different medical specialties (Elstein 1972; Barrows, 1972; and Shulman, 1974). Contrary to earlier beliefs, an outstanding finding made in the three independent investigations is the discovery that physicians generate diagnostic hypothesis early in the patient encounter (Elstein, 1972; Barrows, 1972; Shulman, 1974).

Shulman (1974) pointed out that these hypotheses serve as elements of a conceptual framework which determine the order and the analysis of incoming cues. He describes the conceptual framework as being like a matrix with the cues listed along the vertical axis in the order they are acquired and the hypotheses arranged along the horizontal axis.

As each cue is acquired it is analyzed with respect to each hypothesis. If a cue confirms a hypothesis, Shulman conceptualized that the hypothesis receives a weight of +1, if a cue disconfirms a hypothesis it receives a weight of -1, and if a cue neither confirms nor disconfirms a hypothesis it receives a weight of zero. This conceptual framework of positive/negative ones and zeros serves as a structure to handle the multitude of data that pour from the patient, and guides the physician in the selection of additional cues.

Elstein (1972) claims that the hypotheses are roughly rank ordered according to four principles:

1. Probability: subjective estimates are made of the statistical likelihood that a particular disease is causing the patient's problem. This estimate may closely approximate the population base-rate for a disease.

2. Seriousness: life-threatening or incapacitating conditions are ranked higher than their population base-rate warrants.

3. Treatability: given two equally serious diseases, the treatable one is ranked higher so as not to overlook any treatment which might possibly be helpful.

4. Novelty: some physicians seem to entertain hypotheses which they know are improbable. This strategy seems to keep the physician interested in the case and insures that unlikely avenues are explored.

Elstein (1972) also claims that it is these rank-ordered hypotheses that are systematically tested in a medical work-up.

Barrows (1972), by categorizing and studying the type of question physicians asked, found that two types of questions were used: (1) specific, and (2) general inquiry oriented questions. The specific questions were usually aimed at obtaining detailed items of information, while the general questions were usually aimed at obtaining global items of information. The physician seemed to unconsciously switch back and forth between the two types of questions. When the specific questions were no longer productive or worth pursuing, he unconsciously switched without external evidence to routine general questions. Whenever a positive response came from routine general questions, the physician instantly switched back to specific questions. Barrows also found that physicians used routine general questions whenever they were puzzled or confused. Routine general questions were asked of the patient without exception or concern for positive answers. As the physician was only half listening to the patient's response, he seemed to be re-evaluating his conceptual framework in order to obtain new leads or cues. However, should an unexpected "hit" occur by the patient giving an important answer, the physician picked it up and switched to specific inquiry oriented questions.

Kagan _et_ _al_. (1970) feel that routine general questions insure the clinician did not close prematurely on an obvious diagnosis but kept on carefully searching for general clues that might suggest alternate hypotheses.

Barrows (1972) categorized and studied the types of hypotheses that were generated by physicians. He found that the hypotheses of the experienced clinician were broad and usefully vague. The clinician took several vague hypotheses that popped into his mind early in the interview and allowed them to be shaped by the data derived from his inquiry. Students on the other hand tended to use specific and precise hypotheses unlike the "good" clinician.

Allal (1973) found that hypotheses can be categorized in terms of their relationships with one another. Most often multiple competing hypotheses are formed. That is, a pair (or more) of hypotheses were formulated in such a manner that confirming one implied rejecting the other(s). By rejecting competing hypotheses, it is possible for the physician to transform negative evidence for one hypothesis into a corresponding positive weight for its competitor, thus permitting much more efficient use of our limited human capabilities for information-processing.

Kleinmuntz (1968) demonstrated that data not related to the clinician's mental hypotheses or diagnoses were totally forgotten by the clinician. This finding was substantiated by Barrows (1972). Wason (1968), however, found in a non-medical study that it is extremely difficult for an individual to eliminate a hypothesis as long as there is some confirming evidence. Therefore, a physician with some supporting evidence for a particular hypothesis will find it difficult to reject that hypothesis.

The number of hypotheses that could be held in working memory at any time was found to be clearly limited. Elstein (1972) found that the number of hypotheses entertained at one time seems to be four plus or minus one. This finding was later supported by Shulman (1974). These findings were substantially lower than Miller's (1956) magic number 7 and are in agreement with Simon's (1968) estimate of "five chunk" human mental capacity.

Shulman (1974) found that diagnostic error in medical work is rarely due to an insufficient amount of data. He found that the accuracy of diagnosis is unrelated to the thoroughness of data collected but related to the set of working hypotheses which defines the "problem space" within which the inquiry is conducted. If the problem space is incorrect, then the problem solution will likely be incorrect.

Shulman (1974) compared the diagnostic process of medical students and expert physicians. He found that students accumulate massive amounts of information only to become inundated by its weight and lack of organization. To explain students' diagnostic errors he makes the clear distinction between cue acquisition and cue interpretation. In problem solving, a fact or cue has no meaning per se; its usefulness is derived by its correct use in a particular clinical problem. Thus it is both cue acquisition and interpretation that underlies the diagnostic process.

Elstein (1972) found in his investigations that there was a reasonably high probability that one of the earlier generated hypotheses will become the correct diagnosis.

What justifies the elaborate system of history taking, physical work-up, and laboratory investigations? Hampton et al. (1975) investigated the relative contributions of history-taking, physical examination, laboratory investigation to diagnosis and management of patients. They found that on the average 66 out of 80 patients were correctly diagnosed by 24 clinicians using only referral letters, that seven additional patients were correctly diagnosed using referral letters plus doing physical investigations, and that seven more patients were correctly diagnosed using referral letters, physical examinations, plus laboratory investigations. Thus problem solving generally seemed to occur almost entirely during the interview, while confirmation of the diagnosis seemed to occur within the physical and laboratory investigations. Barrows (1972) felt that there was a direct relationship between the ordered hypotheses and the physical examination; the latter being used to sharpen the former.

Leaper et al. (1974) found that senior clinicians tended to ask fewer questions than their junior counterparts. He also found large individual differences exist between similar clinicians. He stated:

> Each clinician has his own pathway to diagnosis, ...not only does the diagnostic pathway vary from clinician to clinician, but from patient to patient - depending upon such external factors as the difficulty, urgency, and the role which the particular doctor assumes in the management of each particular case. Such an observation explains the great difficulty encountered by our statistical colleagues

> in modelling and delineating in mathematical terms the diagnostic process - for in practice they are attempting to model something which does not exist as a single entity. (p. 152)

2. Clinical Competence

The Funk and Wagnall's Standard College Dictionary defines competent as (1) having sufficient ability; capable, (2) sufficient for the purpose: adequate and (3) having legal qualification; admissible. A physician or student who is clinically competent would therefore have sufficient clinical ability. This definition is vague as the terms "sufficient" and "clinical ability" are not defined. How much is sufficient? What are the abilities? A search through the medical education literature is not very illuminating. Taylor *et al*. (1975) outlined several interdependent abilities upon which clinical competence is based. These abilities are:

1. command of a relevant body of factual knowledge,
2. skills in inter-personal relationships,
3. certain observational and interpretive skills concerned with the gathering of clinical information,
4. a number of decision-making skills collectively referred to as clinical judgement, and
5. certain attitudes which are regarded as desirable in a competent clinician. These include empathy, compassion and altruism.

The National Board of Medical Examiners, with the assistance of the American Institutes of Research, used the critical incident technique developed by Flanagan (1954) "to obtain a definition of clinical competence and skill at the level of the internship, as the young physician with his M.D. degree begins to assume independent responsibility for the care of patients" (Hubbard, 1971). Thirty-three hundred incidents of "good" and "poor" practice were collected, grouped, and classified into the following nine areas:

I. History:

  A. Obtaining information from patient
  B. Obtaining information from other sources
  C. Using judgement

II. Physical Examination:

  A. Performing thorough physical examination
  B. Noting manifest signs
  C. Using appropriate technique

III. Tests and Procedures:

  A. Utilizing appropriate tests and procedures
  B. Modifying test methods correctly
  C. Modifying tests to meet the patient's needs
  D. Interpreting test results

IV. Diagnostic Acumen:

  A. Recognizing causes
  B. Exploring condition thoroughly
  C. Arriving at a reasonable differential diagnosis

V. Treatment:

  A. Instituting the appropriate type of treatment
  B. Deciding on the immediacy of the need for therapy
  C. Judging the appropriate extent of treatment

VI. Judgment and Skill in Implementing Care:

   A. Making necessary preparations
   B. Using correct methods and procedures
   C. Performing manual techniques
   D. Adapting method to special procedure

VII. Continuing Care:

   A. Following patient's progress
   B. Modifying treatment appropriately
   C. Planning effective follow-up care

VIII. Physician-Patient Relationship:

   A. Establishing rapport with the patient
   B. Relieving tensions
   C. Improving patient cooperation

IX. Responsibilities as a Physician:

   A. For the welfare of the patient
   B. For the hospital
   C. For the health of the community
   D. For the medical profession

The nine areas and their subdivision's became the National Board's definition of clinical competence and "constituted a well documented answer to the question of what to test" (Hubbard, 1971). What to test, however, does not answer the question of how to test.

3. Simulations, Techniques Used to Assess Clinical Competence

The use of conventional methods of evaluating medical candidates is often not optimally suited to assess "clinical competence." For example, the oral examination is often used in evaluating a candidate's performance in a clinical situation. However, in a clinical oral at least three sources of variation contribute to the candidate's score; namely the candidate, the examiners, and the patient.

Simulation techniques are often used to reduce the variation due to the examiners and the patient. Bobula and Page (1973) define simulations as follows:

> Reduced to its essence, simulation consists in placing an individual in a realistic setting where he is confronted by a problematic situation that requires a sequence of inquiries, decisions and actions. Each of these activities triggers appropriate feedback which may modify the situation and be used for subsequent decisions about what to do next. The examinee's next action in turn may further modify the problem. Thus a problem evolves through many stages until it is terminated when the individual reaches an acceptable resolution or is faced by unacceptable consequences brought about by his own choices and actions. (p. 1)

Two forms of simulations have evolved in medicine: (1) realistic and (2) abstract simulations. In realistic simulations, the patient/physician setting is a copy of the actual clinical environment. The patient's role is played by an actor or actress and the physician's role is played by the examinee. The patient-physician interaction occurs in a mock-up of the physician's office. This method, although more like the actual physician-patient encounter, has proven to be too expensive in terms of examiner time and costs. Thus the realistic patient simulation is generally replaced by the abstract patient simulation. In the abstract patient simulation, the interaction that might occur between a patient and a physician is duplicated on paper or on a computer terminal. It is with this latter type of simulation that this investigation concerned itself.

The abstract patient simulations are referred to as patient management problems (PMP) and are currently used to evaluate selected components of clinical competence. Skakun (1975) makes this point very clear by stating "It is erroneous to conclude that PMPs measure clinical competence." What they attempt to measure is some aspect of the global construct of "clinical competence" - that aspect resembling such candidate capabilities as problem-solving, clinical judgement, clinical management, and decision-making skills (p. 2). According to Bobula and Page (1973) a simulation may be used to evaluate or study the following component skills:

(1) Skill in determining what sequence to follow in order to solve a problem

(2) Skill in eliciting information or data

(3) Skill in interpreting data

(4) Skill in avoiding unnecessary and wasteful actions (efficiency)

(5) Skill in using a variety of resources, including expert advice

(6) Skill in manipulating a situation to alter it

(7) Skill in monitoring the effects of this manipulation and intervening in reaction to adverse effects

(8) Skill in resolving a problem most effectively (proficiency)

4. Historical Development of Patient Management Problems

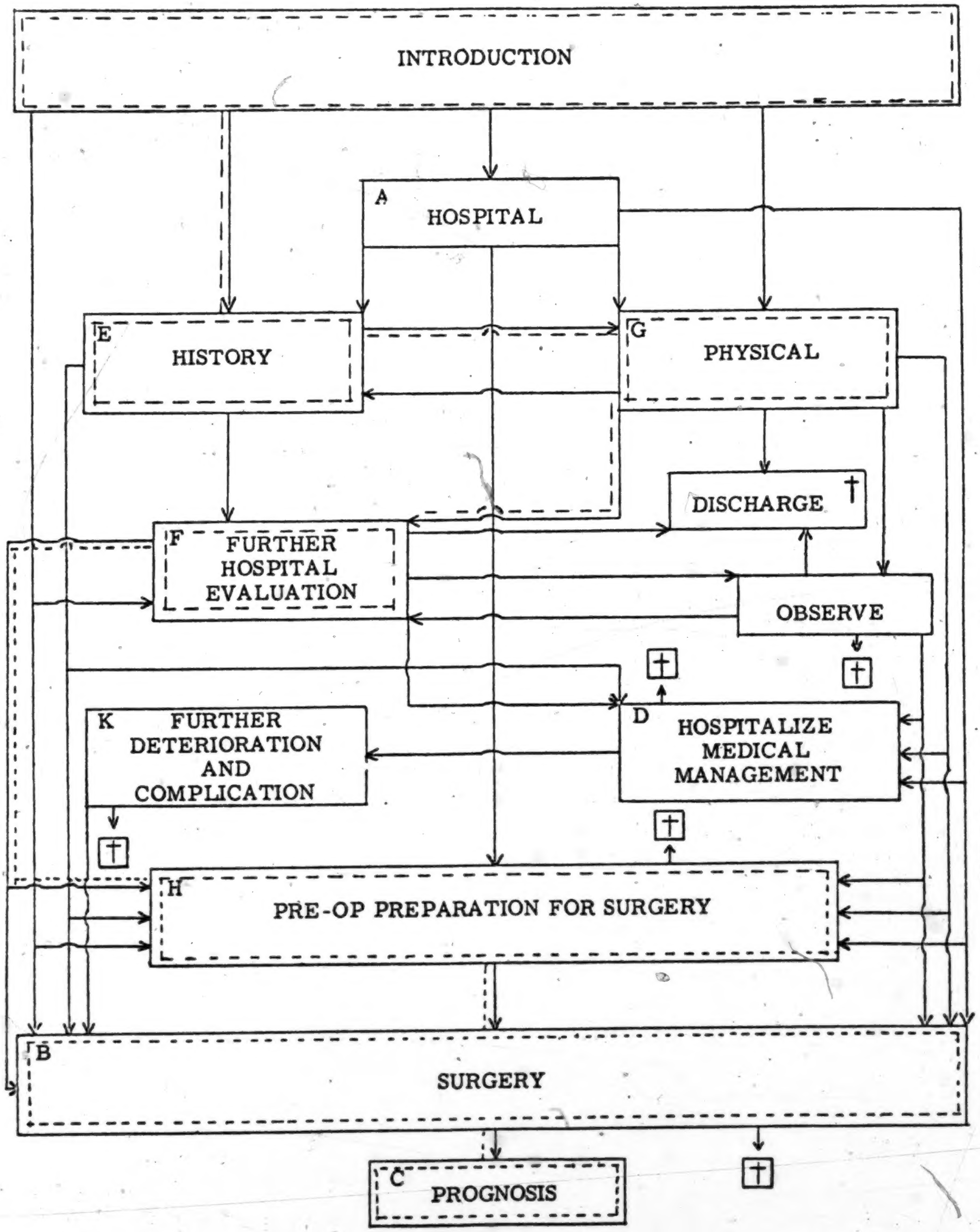
Present day PMPs are derivatives of the Test of Diagnostic Skills (TDS) first introduced by Rimoldi (1955, 1961, 1963). The test consists of cards contained in flat pockets which overlap and are evenly arranged on a display folder. On the top edge of each of these cards a question that the examinee may ask is indicated. These include questions that he may wish to ask of a patient; the manipulative techniques he might wish to use; the diagnostic tests he might order; and so forth. By selecting and looking at the reverse side, the subject gets information that is given in the form of verbal reports, laboratory analysis, x-ray films, etc. For instance for a question like, "Chest x-rays," the answer may be "Both lung fields are normal." The experimenter or the subject writes the number of each item as soon as it is chosen, or, if the cards are perforated, inserts them face down on a pin in the same order in which they are selected. By inspecting the pile of cards the examiner knows both the cards selected and the order of selection. Rimoldi developed the test mainly to estimate how a medical student proceeds when diagnosing a clinical case.

In 1961, the U.S.A. National Board of Medical Examiners became the first licensing agent to utilize the PMP to evaluate clinical competence. They used paper, opaque paint, and an erasure to simulate the clinical encounter. The questions were placed in sections that were linearly arranged. Their linear arrangement of sections became known as the linear PMP model. In this model there is one pathway through the sections. The sections are sequentially arranged and the examinee begins the simulated clinical encounter in section one and proceeds from one section to the next until the last section is completed. In each section, the examinee selects options thought to be relevant. By erasing the opaque layer of paint corresponding to the selected option, the examinee is informed of the consequence of each choice. Step by step the examinees progress linearly through the test selecting information which would lead to a diagnostic decision. A graphical representation of the linear model is outlined in Figure 1.1.

Section I	Opening Introduction to Scene of Patient's Problem
Section II	History
Section III	Diagnosis
Section IV	Physical Investigation
Section V	Laboratory
Section VI	Treatment
Section VII	Prognosis

Figure 1.1 Linear PMP Model

In 1967, the PMPs developed by Rimoldi and the National Board were elaborated on by McGuire and Babbott in an attempt to create an objective, easily administered test that simulates "real" clinical program-solving. They developed the branching model. In the branching model, there is more than one pathway through the sections of the PMP. The many sections are interlinked by a branching device called a "bridge" (McGuire and Babbott, 1967). The student or physician begins the clinical simulation in section one - the opening section. The opening section is generally followed by a bridge which allows the student or physician to select a course of action (i.e., selection of one of the following: hospitalize patient, take brief history, perform emergency treatment, seek consultant's advice, order laboratory tests, perform physical examination). In each section, the student or physician selects options thought to be relevant. By erasing the opaque layer of paint corresponding to the selected option, the student or physician is informed of the consequence of each choice. Step by step the examinee branches through the test selecting information that would result in the best health care for the patient. In comparing the linear and branching models, the branching model goes beyond the linear model by allowing the student or physician to select one of many courses of action. For a graphic example of a branching model see Figure 1.2 which was obtained from the handbook composed by the University of Illinois Evaluation Unit (1967).


INTRODUCTION
A
HOSPITAL
E
HISTORY
G
PHYSICAL
DISCHARGE
†
F
FURTHER HOSPITAL EVALUATION
OBSERVE
K
FURTHER DETERIORATION AND COMPLICATION
D
HOSPITALIZE MEDICAL MANAGEMENT
H
PRE-OP PREPARATION FOR SURGERY
B
SURGERY
C
PROGNOSIS

Dotted lines = Optimal route

☩ = Death of patient

Figure 1.2 Diagram of a Branching PMP.

In 1971 Helfer and Slater developed an instrument that they called the diagnostic management problem (DMP). The DMP is a slight modification of the test developed by Rimoldi (1961). The examinee is presented with a deck of cards, told the setting in which he is working, given a brief abstract of the case, and provided with an index sheet which itemizes the type of information available on each numbered card. Instead of looking at the top edge of each card for a question that might be asked, (i.e., TDS), the examinee looks at the index sheet. The major difference between TDS and the DMP are the scores that are calculated to describe the diagnostic processes. This difference will be discussed in the next chapter.

In 1970 computers were being used at all major universities and medical educators began using their potential to simulate clinical encounters. Harless *et al*. (1971) developed the simulated patient encounter known as the Computer-Aided Simulation of the clinical Encounter (CASE). In CASE, the computer begins the session by presenting on a computer terminal a brief description of the patient's problem. The student then is "free" to query the computer in natural language regarding any aspect of the patient's medical problem. The student is allowed to use his own problem-solving style and his own method of inquiry. There are no cues, such as a dictionary of acceptable questions, to influence the student's path of inquiry and no artificial language to restrict his interaction with CASE. The interaction is, however limited by the sophistication of the computer algorithm used to analyze requested information.

In 1973, the R.S. McLaughlin Examination and Research Centre began experimenting with PMPs for the assessment of clinical competence. The first attempt was a joint project with the National Board of Medical Examiners. The linear model of the National Board was utilized to examine Paediatric fellowship candidates. However, in 1974, the McLaughlin Examination and Research Centre developed linear and branching PMPs which were administered across Canada on computer terminals. These computer-presented patient management problems (CPMP) were similar in form to the linear and branching models discussed earlier, but their scoring techniques differed. The scoring procedure used by the R.S. McLaughlin Examination and Research Centre will be discussed in the next chapter.

In 1973, Friedman constructed a computer patient model which utilized many of the aspects of previous systems but which added a time axis, so that the length, cost, availability and effect of tests or procedures became, as they are in a real hospital situation, an important part of case work-up. After presenting an opening scene which briefly described the patient's condition, the physician is free to request any test he desires in any order he desires, and he may make a diagnosis at any time during the encounter. Friedman compared the performance of medical students to practising physicians and found that medical students had the longest lapsed time, spent the most money, and kept their patients in the hospital for the longest period of time.

In 1974, Berner et al. described a pencil and paper instrument they had developed to evaluate clinical problem-solving. Their format was designed to simulate reality, be convenient to administer to large groups, be easily and objectively scored, and in addition minimize the effect of cueing. It is a combination of Weeds (1969) problem-oriented approach to clinical thinking and record-keeping, Soloman's sequential Management Problem (SMP), and the PMP developed at the Medical College at the University of Illinois. The main advantage of this instrument over the other pencil and paper PMPs is that it minimizes cueing and enables the examiner to determine why specific options were chosen by the examinees.

In summary a variety of PMPs have been developed since the initial work of Rimoldi in 1955. However, with the variety of PMPs came a variety of scoring procedures. These scoring procedures are briefly outlined in the following section.

5. Scoring Procedures

A key must be first developed in order to score examinee selections. There are two methods for calculating examinee scores. The first is based upon decision theory, probabilities, and Baysian statistics. The second is a linear model which involves the summing of assigned weights. Both methods of scoring have their advantages and disadvantages.

Shulman (1972) argues against the use of probability statements and Baysian statistics. He claims that physicians use probabilities, if at all, only in a most imprecise, intuitive fashion and their subsequent revisions of hypotheses in light of new data do not conform to Bayes' Theorem. He instead supports the use of the linear scoring model.

The linear scoring model is the simplest of the two methods and is used with all PMPs described in the previous section. For this latter reason, this investigation will restrict itself to the investigation of the different applications of the linear model.

The basic linear model may be represented as follows:

$$X_j = W_{11}S_{11j} + \ldots + W_{pn}S_{pnj}$$

where X_j is the score given to the jth examinee,

W_{pn} is the weight assigned to the pth decision in the nth section, and

S_{pnj} is the selection made on the pth decision of the nth section by the jth examinee (*i.e.*, 1 = selected and 0 = not selected).

Using the linear model generally involves two steps. First, it is necessary to categorize decisions as either correct or incorrect for the solution of the patient's problem.

Naturally the correct decisions are those that should be selected and the incorrect decisions are those that should be avoided. Secondly, the categories must be weighted. The weight assigned generally reflects the appropriateness or inappropriateness of the decision with respect to the optimal health care of the patient.

In all PMPs described in section 4, expert physicians are used to categorize and weight decisions as either correct or incorrect. The experts used in each testing situation may vary depending upon the candidates examined.

In summary, research studies indicated that the following characteristics were included in the diagnostic process:

(1) examination and evaluation of presenting signs and symptoms,

(2) early formulation of global diagnoses or hypotheses,

(3) the use of hypotheses to guide information gathering,

(4) the restructuring of the hypotheses on the basis of the new information, and

(5) the establishment of a diagnosis, and the selection of a treatment based on the diagnosis made.

A variety of forms of simulated patient management problems (PMPs) have been developed to evaluate the above underlying characteristics of the diagnostic process and a variety of procedures have been used to develop scoring keys. Those procedures used by the National Board of Medical Examiners, the College of Medicine, University of Illinois, and the R.S. McLaughlin Examination and Research Centre will be examined in detail in the following chapter.

CHAPTER III

CURRENT SCORING PROCEDURES

In order to examine the various scoring procedures currently used, a general classification system will be presented. In a patient management problem, the key represents the categorization of acceptable and unacceptable clinical decisions, and the number of marks awarded to each decision made. Basically there are three methods for categorizing decisions: (1) group consensus, (2) individual judgements, and (3) computer performance. In group consensus a panel of experts collectively categorize each decision; in individual judgement, each expert independently classifies each decision; and, in computer performance, each expert independently solves the simulated patient's problems at a computer terminal. In the latter method expert selections are used to categorize decisions as being either correct or incorrect.

Scoring keys indicate the number of marks awarded for correct and incorrect decisions. Basically two types of weights are assigned; (1) constant and (2) differential. In a constant weighting system an equal number of marks are awarded for each correct and incorrect decision (e.g., correct decision #101 = +5, incorrect decision #102 = -5). In a differential weighting system an unequal number of marks are awarded for each correct and incorrect decision (*e.g.*, correct decision #11 = +6, correct decision #21 = +3, incorrect decision #31 = -1, incorrect decision #41 = -4). In both the constant and differential weighting systems, no marks are either lost or gained for selecting decisions that are categorized as neither correct nor incorrect.

There may be single or multiple keys depending upon whether there are valid variations in expert judgements and/or performances. If a single key is used then there is one acceptable and unacceptable set of decisions to be made. If multiple keys are used then there is more than one set of decisions to be made.

Lastly, scoring techniques may vary in the method of awarding marks. Basically, there are two methods currently used: (1) sum and (2) true/false (T/F). In the sum method marks are awarded for selecting correct options and subtracted for selecting incorrect options. In the T/F method marks are awarded for selecting correct options and for not selecting incorrect options; no marks are subtracted.

In summary, scoring procedures can be classified by (1) method of categorizing options - group consensus, expert judgements, and expert performance; (2) assignment of weights - constant and differential; (3) number of keys - single or multiple; and (4) the method of awarding marks - sum and T/F. In order to describe the scoring procedures currently used, each procedure will be defined using the above classification system.

1. Scoring Procedures Currently Used

A. Group consensus, constant weight, single key, and T/F method.

The National Board of Medical Examiners used a pencil and paper management problem to test student's clinical problem-solving skills as part of a certification examination for obtaining a Doctorate of Medicine (MD) degree. The following scoring technique is described by J.P. Hubbard (1971, pp. 47-48).

> The scoring of patient management problems gives credit for correct decisions and penalties for sins of omission and commission. Each of the several hundred choices or courses of action, offered in the test is classified in one of three categories: (1) it must be done for the well-being of the patient; (2) it should definitely not be done, and if done, would be a serious error in judgment that might be harmful to the patient; and (3) it is relatively unimportant, i.e., a procedure that might or might not be done, depending upon local conditions and customs. Each examinee is given a "handicap score" equal to a total number of items coded as definitely incorrect. Each time the examinee selects an incorrect choice, one point is subtracted from this score; each time he selects a correct choice, one point is added. Thus, his total score on this test is the number of correct decisions he has made, i.e., the number of indicated procedures he has selected plus the number of incorrect procedures that he has avoided. The choices in the equivocal middle ground receive no score.
>
> The programmed testing method is quite different from the usual multiple-choice technique in which the candidate is offered a number of choices and instructed to select one best response. Here, he is offered a number of choices and required to use his best judgment in selecting all those, and only those, he considers important for the management of the patient. Usually, as in a practical situation on a hospital ward, he recognizes a number of actions that should definitely be done and other actions that should definitely not be done. His responses are therefore interrelated. If he is on the right track,

> he makes a number of correct decisions from among the available choices; then, by his erasures, he gains the information necessary for the proper management of the patient in the next problem in the next set of choices. If he starts off on the wrong track in this programmed test, he may compound his mistakes as he proceeds and become increasingly dismayed as he learns from his erasures the error of his ways. If he discovers that he is on the wrong track, however, he has a chance to change his course and to make additional choices, although he cannot undo the errors that he has already committed - again a situation rather true to life.
>
> Since in this testing technique, as in the use of the more traditional type of multiple-choice examinations a panel of experts has determined the rightness or wrongness of each choice or course of action offered to the examinee, accurate and detailed statistical analyses are equally applicable.

In summary, the above scoring procedure uses group consensus to categorize options,a constant of +1 and -1 to respectively weight correct and incorrect decisions, a single key to score examinee selections, and the true/false method to calculate examinee scores. (For elaboration see pg. 48).

B. Group consensus, differential weights, single key, and sum method.

At the University of Illinois, College of Medicine, Christine McGuire pioneered the development of the branching PMP model (McGuire and Babbott, 1967; McGuire and Soloman, 1971). Her developmental work has resulted in an increased use of patient simulation by medical educators in North America. The scoring technique developed by the University of Illinois is described below:

Using a group of experts each option in a problem is placed in one of the following categories:

++ Category: Choices which are CLEARLY INDICATED and IMPORTANT in the care of THIS patient at THIS stage in the workup or management;

+ Category: Choices which are CLEARLY INDICATED but of a more ROUTINE nature, *i.e.*, should be selected but are not of special significance in the care of THIS patient at THIS stage;

0 Category: Choices which are OPTIONAL, *i.e.*, the probability that they will be helpful for THIS stage is fairly remote or quite debatable;

- Category: Choices which are clearly NOT INDICATED though NOT HARMFUL in the management of THIS patient at THIS stage;

-- Category: Choices which are clearly CONTRA-INDICATED (*i.e.*, are definitely harmful or carry an unjustifiable high cost in terms of risk, pain or money) in the care of THIS patient at THIS stage.

In addition to the above categories, a further classification is made on the ++ and -- categories. Some of the options in these categories are further divided into two additional categories (*i.e.*, +++, and ++++; or ---, and ----). Further division depends upon the degree of urgency or importance of either selecting or avoiding the particular option.

Once each option is categorized, as described above, weights are assigned that reflect the option's relative harm or help in the management of the patient. While any set of weights can be employed, the following weights are commonly used:

Weight	Category
+ 16 points	For any option in the "++++" category
+ 8 points	For any option in the "+++" category
+ 4 points	For any option in the "++" category
+ 2 points	For any option in the "+" category
0 points	For any option in the "0" category
- 2 points	For any option in the "-" category
- 4 points	For any option in the "--" category
- 8 points	For any option in the "---" category
- 16 points	For any option in the "----" category

Each student/physician's total score is calculated by summing the weights corresponding to correct decisions, and subtracting weights corresponding to incorrect decisions. The above scoring technique is described in a handbook titled, "Materials for the Evaluation of Medical Performance in Medicine, 1967."

In summary, the above scoring procedure uses, group consensus to categorize decisions into one of nine categories, the categories to differentially weight decisions, one key to score examinee selections, and the sum method of calculating examinee scores.

C. Group performance, constant weight, single key, and sum method.

The R.S. McLaughlin Examination and Research Centre mainly use the linear CPMP model. In the 1974 Paediatrics Examination, four CPMPs were used. Three of the CPMPs required two types of responses - (1) select, and (2) select and rank order. Due to the two types of responses, their scoring techniques were considerably more complex. For simplicity and comparative purposes, only the select responses will be described.

Fifteen practising paediatricians who had not seen the computer patient management problem before took the examination. Their decisions on each option were recorded and a scoring key was developed based on the number of experts selecting each option. The following criteria were used to categorize the options.

(1) correct (+) if 8 or more experts selected the option,

(2) neither correct nor incorrect (0) if 7 or fewer experts selected the option, and

(3) incorrect (-) if no experts selected the option.

A constant weight of 5 was assigned to the above categories 1 and 3; +5 marks were awarded to candidates for selecting correct options, and 5 marks were subtracted for selecting incorrect options. No marks were added or subtracted for options categorized as neither correct nor incorrect. An example of the scoring key using the above procedure is presented in Table 3.1.

TABLE 3.1

Example of Scoring Key Based on Group Performance, Constant Weight, Single Key, and Sum Method

	OPTION									
	101	102	103	104	105	106	107	108	109	110
Number of experts selecting	12	5	7	0	9	1	4	3	6	0
Key	+5	0	0	-5	+5	0	0	0	0	-5

In conclusion, the above key uses the selections of expert problem solvers to categorize options, a constant of +5 and -5 to respectively weight correct and incorrect decisions, one key to score examinee selections, and the sum method to calculate examinee scores.

D. Group consensus and performance, differential weights, single key, and sum method.

In a Meningitis Management Problem developed by the R.S. McLaughlin Examination and Research Centre, the committee of expert paediatricians who designed the problem classified each option into three categories: (1) correct (+), should be selected; (2) neither correct nor incorrect (0), may be selected; and (3) incorrect (-), should not be selected.

In addition, eleven expert paediatricians who had not seen the problem before, took the examination. Their performance was used to weight each of the (+) and (-) options. The (+) options were weighted using the following formula:

$$+Wi\% = \frac{\sum_{j=1}^{k} N_{ij}}{\sum_{i=1}^{m} \sum_{j=1}^{k} N_{ij}} \times 100 \qquad (3.1)$$

where $+Wi\%$ is the weight for (+) option i expressed as a percentage.

N_{ij} is the decision (1 = selected and 0 = not selected) for expert j on (+) option i.

$\sum_{j=1}^{k} N_{ij}$ is the number of k experts who selected (+) option i.

$\sum_{i=1}^{m} \sum_{j=1}^{k} N_{ij}$ is the total number of selections made by k experts on m (+) options.

Thus, (+) Wi% is proportional to the number of experts selecting (+) option i compared to the total number of selections made by k experts on m (+) options.

The (-) options were weighted in a similar manner using the following formula:

$$-W_{i\%} = \frac{K - \sum_{j=1}^{J} 0_{ij}}{MK - \sum_{i=1}^{,m} \sum_{j=1}^{k} 0_{ij}} \text{ X } 100$$

where -Wi% is the weight assigned to (-) option i expressed as a percentage

0_{ij} is the decision (1 = selected, and 0 = not selected) for expert j on (-) option i.

$K - \sum_{j=1}^{J} 0_{ij}$ is the number of k experts not selecting (-) option i.

$MK - \sum_{i=1}^{m} \sum_{j=1}^{k} 0_{ij}$ is the total number of k experts not selecting m (-) options.

Thus, (-)Wi% is proportional to the number of experts not selecting (-) option i compared to the total number of k experts not selecting m (-) options.

As the (-) and (+) option weights are expressed as proportions of a total, their sums would equal -100% and +100% respectively (*i.e.*, Σ(-)Wi% = (-)100%; Σ(+)Wi = (+)100%). Therefore if all options were circled, the total score would equal (+)100% + (-)100% = 0%. See Appendix B for categorization and weights that were assigned to each option in the Meningitis problem.

In summary, the above scoring procedure uses group consensus to categorize options, group performance data to differentially weight options, a single key to score examinee selections, and the sum method to calculate examinee scores.

2. Comparison of Current Scoring Procedures

A summary of the four current scoring techniques is presented in Table 3.2. A comparison of the scoring procedures reveals similarities and differences.

A. Linear versus branching.

There is a great deal of controversy over the pros and cons of using a linear or branching model for assessing problem-solving skills. Hubbard (1971) writes:

> Although the branching program may seem attractive, and has been introduced by McGuire (1963) as a modification of the PMP test, the National Board has held to the linear method to assure that each examinee is tested with essentially the same

TABLE 3.2

Categorization of Proposed Scoring Procedures

	I National Boards	II University of Illinois	III R.S. McLaughlin Examination and Research Centre	IV R.S. McLaughlin Examination and Research Centre
PMP model	linear	branching	linear	linear
Responses	select	select	select and/or scale	select
Categorization of options	expert consensus	expert consensus	group performance	group consensus and performance
Weights	constant (i.e., +1, 0, -1)	differential (i.e., +16 to -16)	constant (i.e., +5, 0, -5)	differential (i.e., +6.9% to -3.8%)
Key	single	single	single	single

> examination. When unlimited branching is permitted, two different examinees may take totally different approaches to the clinical situation and follow different pathways to the solution of the problem. In this case there is no accurate way to evaluate the two examinees except in terms of whether or not they ultimately solved the problem (i.e., gave the "correct" final diagnosis) regardless of what they had done (or not done) for the patient in the interim.

Hubbard's criticism is based on the difficulty of accurately assessing clinical decisions when there are unlimited branches leading to the correct problem solution.

McGuire (1967) criticizes the linear model as being superficial:

> The earlier tests are in linear form: each examinee is confronted with the same problem, which remains identical throughout for all respondents; thus a premium is necessarily placed on efficiency in reaching a single, correct solution or on the appropriateness of each decision independently. In contrast, the branching problems . . . require the subject to make revealing choices from an almost unlimited number of broad strategic routes, several of which may lead to an acceptable result. (p. 10)

Irrespective of the pros and cons of the current linear or branching PMP models, any of the scoring techniques could be used on either model. In other words using expert consensus and/or performance, constant or differential weights, T/F or sum method, or single or multiple keys remains independent of whether the linear or branching model is used. The applicability of linear vs. branching model will be examined in the present study in combination with different scoring procedures.

B. Expert consensus and/or performance

One of the major problems in developing a key is to decide whether it is more appropriate to use expert consensus and/or expert performance to categorize options. Currently there are no guidelines or research results to assist in making a decision. Each procedure has its advantages and disadvantages.

In group consensus, experts treat the test as a problem with a known answer and categorize the options accordingly. Knowing the answer however does not eliminate disagreements among experts. One expert may consider an option as highly necessary for the particular patient, another equally eminent expert may view the option as necessary, unimportant, or even a waste of money and time. The resolution of the disagreements is dependent upon the structure and processes within the group.

In expert performance, experts begin the test as a problem with an unknown answer. Decisions made reflect the problem-solving behavior of the experts. As in group consensus, options selected by experts while problem-solving also vary among experts. The resolution of the disagreements is dependent upon the mathematical or statistical procedure used to summarize the data.

Categorizing and weighting options while knowing the answers and being thoroughly familiar with the patient simulation is not the same as the mental process faced by experts who select options while problem-solving a patient simulation seen for the first time. The saying that "hindsight is better than foresight" is certainly applicable here. Knowing the end result would certainly allow the direct/optimal course of action to be plotted; but this course of action may not be the same as defined in an optimal problem-solving strategy.

In order to develop a key which outlines the "optimal problem-solving" strategy, group performance data are used; however, to develop a key which outlines the direct/optimal solution, group consensus data are used. Can these two procedures be combined?

The question of whether to use group consensus or expert performance seems to have resulted in the use of both as illustrated by the Meningitis Problem described in example D of the previous section (p. 36). Whether the advantages of either group consensus or group performance outweight their disadvantages is unknown to the author, but using both does eliminate having to choose one over the other.

Using both group consensus to categorize options and expert performance to calculate differential weights by formulae 3.1 and 3.2, may however produce a key which does not reflect the perceptions of either group. An examination of the key derived for the Meningitis problem clearly indicates discrepancies between options categorized by group consensus and options selected by expert performance.

In the Meningitis problem, 23 out of 57 options were categorized as correct; these options were perceived by the panel of experts who constructed the PMP to be "necessary for the optimal health care of the patient." Yet, many of the options categorized as "correct" were not selected by the eleven expert paediatrician problem-solvers. If all eleven expert problem-solvers had selected all 23 options categorized as "correct," there would have been 23 X 11 = 253 correct selections made. However, the expert problem-solvers selected only 160 out of the possible 253 selections (i.e., 63.25% of the (+) options were selected). Ninety-three (36.75%) "correct" options were not selected by the expert problem solving group. This failure to select correct options represents a high error due to omission.

Thirty out of 57 options were categorized as incorrect; these options were perceived by the panel of experts who constructed the PMP to be "detrimental for the optimal health care of the patient." Many of the options categorized as "incorrect" were not selected by the eleven expert paediatrician problem-solvers. If all eleven expert problem-solvers had not selected all 31 options categorized as "incorrect," there would have been 31 X 11 = 341 correct decisions made. The expert problem-solvers did not select 286 out of the possible 341 options categorized as incorrect (i.e., 83.87% of the "incorrect" options were not selected). Fifty-five (16.13%) "incorrect" options were selected by the expert problem-solving group. This selection of incorrect options represents a relatively low error due to commission.

In summary, it would be possible to describe the errors of the expert group as being higher due to omission than due to commission. However, this behavior pattern is not reflected by the differential weights assigned to each option using formulae 3.1 and 3.2.

If all eleven specialists selected an option categorized as correct it received a weight of 6.9% (i.e., $6.9\% = \frac{11}{160} \times 100$). Yet if all eleven specialists did not select an option categorized as incorrect it received a weight of only 3.8% (i.e., $3.8\% = \frac{11}{286} \times 100$). Plotting the weights assigned by the number of specialists respectively selecting and avoiding correct and incorrect options would reveal a difference in slopes of the two lines. See Figure 3.1.

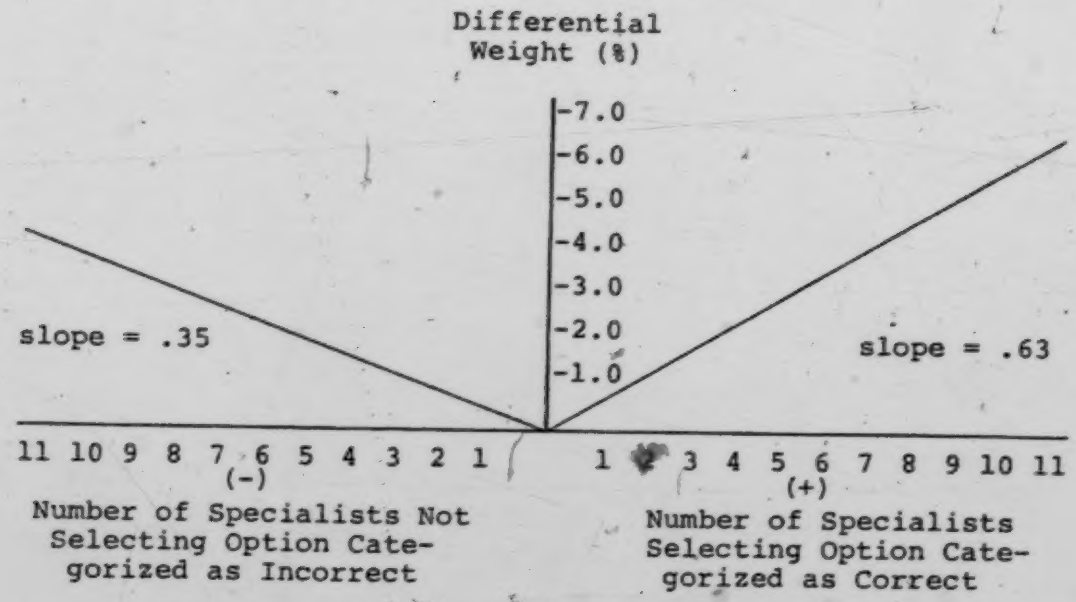


Figure 3.1. Difference in Slope Between (+) and (-) Weights Assigned in Meningitis Problem.

The slope for the (-) options is approximately one-half the slope of the (+) option. Thus fewer marks are lost for selecting "incorrect" options than for not selecting "correct" options. This relationship is opposite to the behavior pattern of the specialists. Collectively most specialists did not select "incorrect" options. Yet an examinee would have fewer marks taken away for selecting "incorrect" options and receive higher marks for selecting "correct" options. The assignment of weights would be biased in favor of the "non cautious, over-responding" examinee who tended to guess, and would be biased against the "cautious, under-responding" candidate who tended not to guess.

In conclusion, in the group consensus procedure of developing a scoring key, differences in expert perceptions are resolved through group discussion; and in the expert performance procedure, differences in expert perceptions are reflected by the weights applied to the various options. The cognitive task set for the two groups differ. For group consensus, the task is to categorize and weight options based upon knowing the results of selecting each decision. For expert performance, the task is to select options which will lead to the optimal solution of the simulated patient's problem. Combining both procedures may produce a key that does not reflect either group's consistent perceptions. The effect on examinee scores of using group consensus and/or expert performance to develop a key will be studied in this investigation.

C. Constant or differential weights

A weighting system is still applied whether constant or differential weights are used. The important question that must be answered is: What weights should be used and what is the best method of obtaining them? The National Board and the University of Illinois, College of Medicine, both use group consensus to weight and categorize options. However, the National Board assigns a constant weight of (+) 1 and (-) 1 to correct and incorrect decisions, while the University of Illinois generally assigns differential weights of (+) 16 to (-) 16. The main difference between the two weighting systems is that the National Board does not differentiate between levels of "appropriateness" among the decisions to be made in solving the simulated patient's problem while the University of Illinois does.

The R.S. McLaughlin Examination and Research Centre uses expert performance to assign both constant and differential weights. There may be a disadvantage in using group performance to differentially weight options. If two options were both selected or avoided by all specialists taking the examination, both options could be equally weighted. The equal weights however may not reflect the "appropriateness" or inappropriateness of the decisions in solving the simulated patient's problem. One option may be very highly inappropriate (i.e., administer a drug that would lead to the patient's death) and be avoided by all experts. Another option may be a routine investigation (i.e., order a complete blood count), and be selected by all experts. Using expert performance, however, to weight the above two options, will result in an equal number of marks awarded to an examinee who selects one and avoids the other.

Lord and Novick (1968) point out that in evaluating any weighting system, it is necessary to show that the system adds more relevant ability variation than error variation. The amount of residual information can be recovered by differential weighting is subject to question, and more importantly to experimental study (p. 134). The effect that constant and differential weights have upon examinee scores will be studied in this investigation.

D. True/False or Sum method.

The National Board assigns each examinee a "handicap" score equal to the total number of items coded as definitely incorrect. Each time the examinee selects an incorrect option, one point is subtracted from the "handicap" score; each time the examinee selects a correct option, one point is added (Hubbard, 1971). A closer examination of this technique would show that identical scores are produced by scoring each option as either true or false. This fact is illustrated by Figure 3.2.Thus it may be concluded that the National Board of Examiners uses the T/F method of calculating examinee scores; other institutions however use the sum method.

What are the similarities and differences between the T/F and sum method? In order to compare scores calculated by the two methods, suppose a PMP containing 100 "correct" options and 200 "incorrect" options was administered to a group of examinees and their scores were calculated by both methods. What observations would be made if each correct option was assigned a weight of +1 and each incorrect option a weight of -1?

The maximum and minimum scores would be 300 and 0 for the T/F method, and 100 and -200 for the sum method. The range between the maximum and minimum would be equal to 300 for both the T/F and sum methods. These two scales are summarized in Figure 3.3. If a plot were made of the examinee scores calculated by the T/F and sum methods, the points

Options	Options Classification + = correct 0 = neither - = incorrect	Candidate's selection 1 = selected 0 = did not select	T/F Scoring Method	National Board's Scoring Method. Handicap score = 4
Section I				
101	+	1	1	+1
102	-	0	1	0
103	+	0	0	0
104	+	1	1	+1
105	0	1	0	0
106	-	0	1	0
Section II				
201	+	1	1	+1
202	-	0	1	0
203	-	1	0	-1
204	0	0	0	0
205	+	1	1	+1
206	+	1	1	+1
207	0	1	0	0
208	+	1	1	+1
TOTAL	7 +s 4 -s	9 selected 5 not selected	9	4+6-1 = 9

Figure 3.2 The Total Score Using the National Board or the T/F Method Produces a Score of 9.

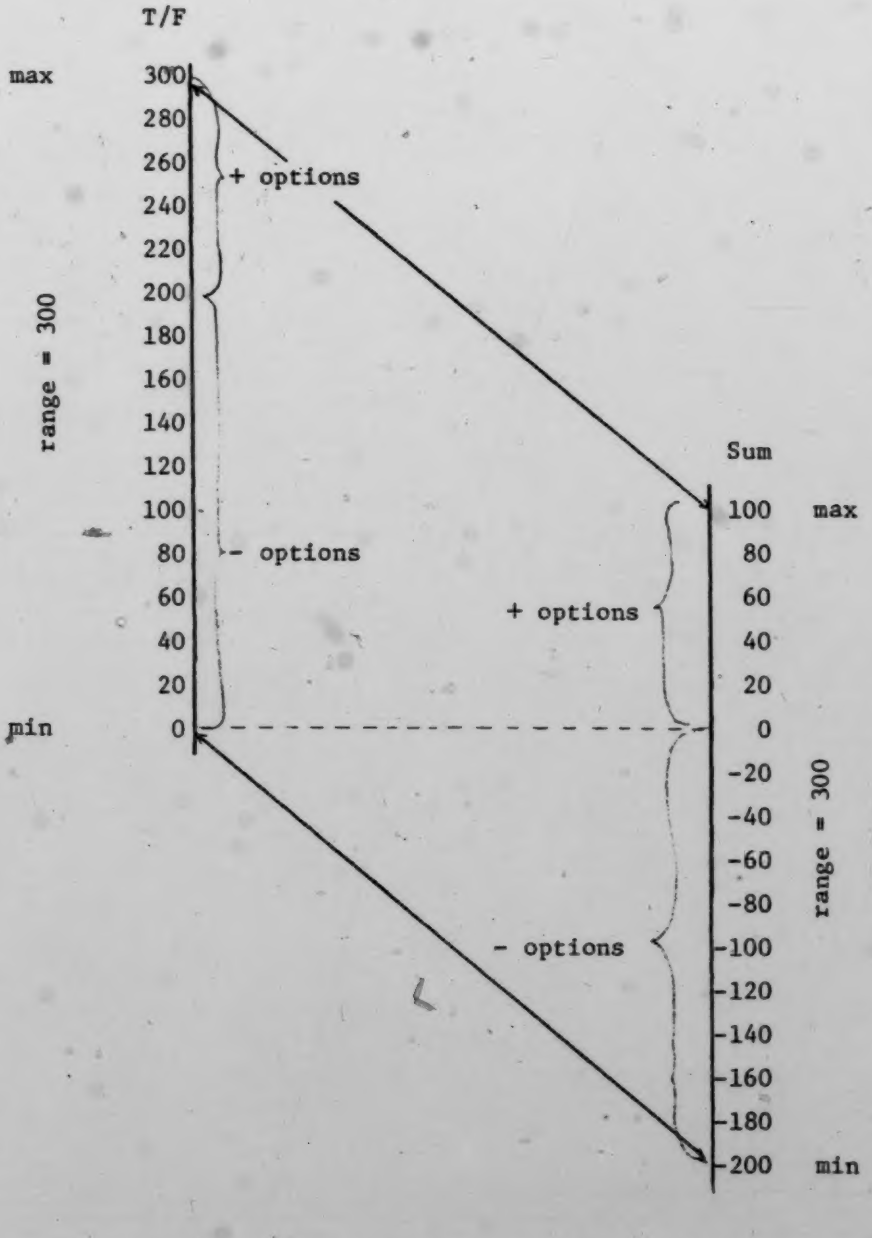

T/F
max
300
280
260
240
220
200
180
160
140
120
100
80
60
40
20
0
min
+ options
- options
range = 300
Sum
100
max
80
60
40
20
0
-20
-40
-60
-80
-100
-120
-140
-160
-180
-200
min
+ options
- options
range = 300

Figure 3.3. Comparison of T/F and Sum Scores.

would fall on a straight line (see Figure 3.4).

The two sets of scores would be perfectly correlated with a slope equal to 1.0 This linear relationship may be summarized as follows:

$$Y_i = X_i + 200 \qquad (3.3)$$

where Y_i is the ith examinee's score calculated by the T/F method,

X_i is the ith examinee's score calculated by the sum method, and

200 is the difference in maximum values between the T/F and sum methods (i.e., 300 - 100).

If examinee percentage scores were calculated for the T/F and sum methods, and plotted, the points would also fall on a straight line (see Figure 3.5).

The two sets of % scores are perfectly correlated ($r = 1.0$) with a slope equal to 1/3. This linear relationship is summarized as follows:

$$\begin{aligned} \%Y_i &= 1/3(\%X_i + 200) \\ &= .33\ \%X_i + 66.67 \end{aligned} \qquad (3.4)$$

where $\%Y_i$ is the % score for the ith examinee calculated by the T/F method, and $\%X_i$ is the % score for the ith examinee calculated by the sum method,

1/3 is the ratio of the maximum scores for the sum and T/F methods (i.e., $\frac{100}{300}$), and

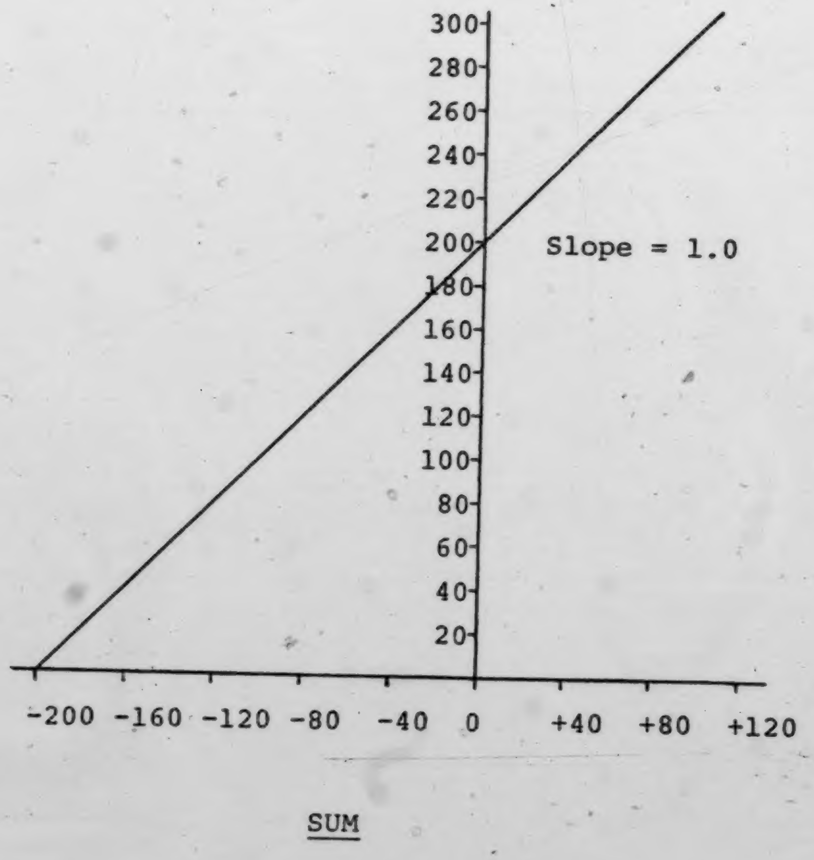


Figure 3.4 Linear Relationship Between Scores Calculated Using True/False and Sum Methods

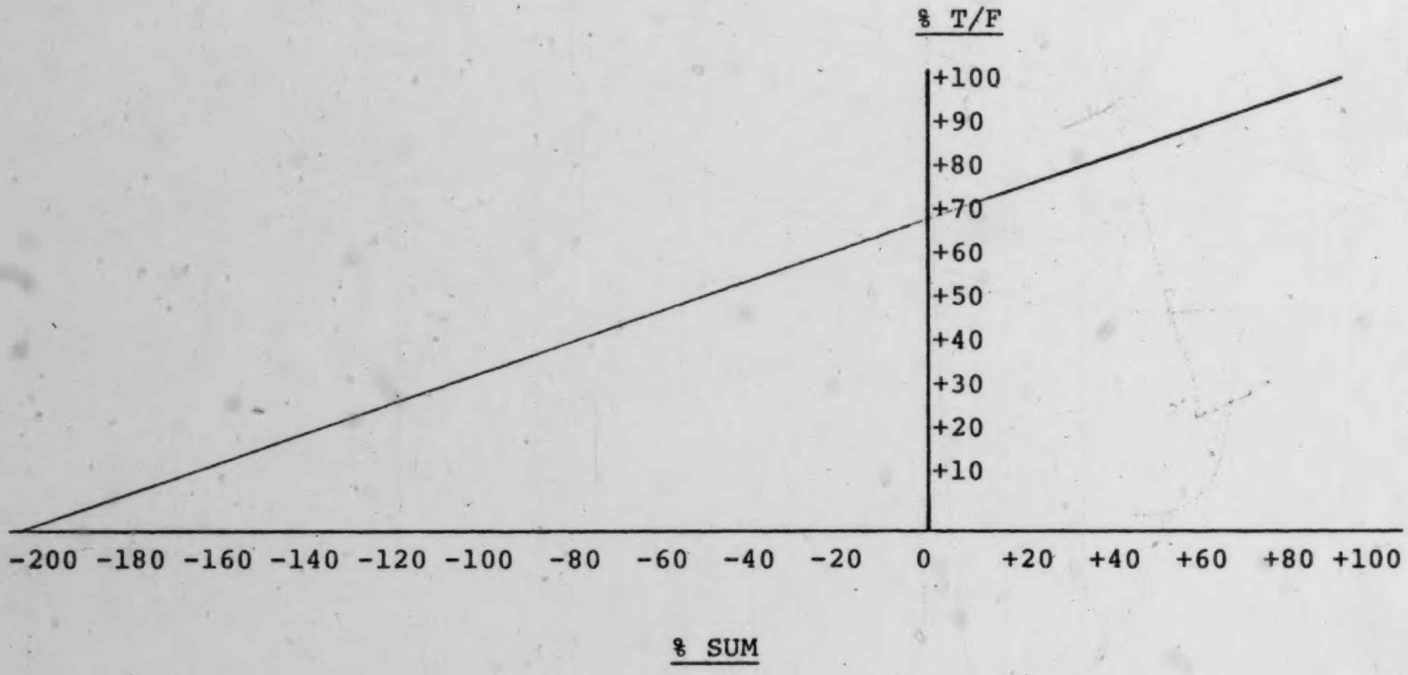


Figure 3.5. Linear Relationship Between % Scores Calculated Using Sum and True/False Method.

200 is the difference of the maximum scores for
the sum and T/F methods (i.e., 300-100).

The length of the % T/F and % sum scales respectively equal 100% and 300%. If the length of % T/F and % sum scales for other simulations were examined, the length of the % T/F scale would always equal 100%, but the length of the % sum scale would vary according to the emphasis placed on - and + options. For example, if 50 and 150 marks were respectively assigned to - and + options, the length of the % T/F scale would equal 100% (i.e., 25% for - options and 75% for + options) but the length of the % sum scale would equal 133% (33% for - options and 100% for + options). The proportion of marks allocated to - and + options is the same for both scales (i.e., 1/4 and 3/4 respectively for - and + options).

Care should be employed when interpreting % scores calculated by either method. Percentage scores calculated by the T/F method should be interpreted as reflecting the percentage of marks gained by respectively selecting and not selecting correct and incorrect options; while % scores calculated by the sum method should be interpreted as reflecting the percentage of marks gained by selecting correct and lost by selecting incorrect options.

It is inappropriate to make direct comparisons of scores if the scales are different. For this reason, it is inappropriate to compare % T/F and % sum scores or even two % sum scores since their scales differ. It is, however, appropriate to compare % T/F scores since their scales are equal. Thus the T/F method will be used to calculate student scores on each simulation.

In conclusion, the raw and percentage scores calculated by the T/F and sum methods correlate perfectly. Care should be employed when interpreting and comparing % scores calculated by either method. If comparisons between scores are to be made, the scores must be on the same scale which is perhaps easiest accomplished by using the T/F scoring method.

E. Single or multiple keys.

All current methods use a single key which reflects the optimal decision of the "average" expert. Although the scores produced using a single key may be straight forward they may lead to false interpretation in that the results may not accurately describe the pattern of responses of either the experts or the candidates. A more appropriate method may be to use individual or subgroups of expert judgements and/or performances to establish more than one key. Thus instead of matching each candidate against an "average" expert and/or optimal problem-solving strategy each candidate would be matched to each expert's or homogeneous subgroups of experts' judgements and/or "problem solving" strategies.

The underlying rational for using a single key is based upon the assumption that the probability of error in judgement and/or performance is randomly distributed among the experts; and, that the average judgements and/or performance is the best estimate of the "optimal" decisions within the patient management problem. A single key also assumes that there is only one "optimal" set of decisions and that variation is due to error in the expert's opinion and not to the nature of the problem.

At the other extreme, if the judgements or performance of each expert formed a key, there might be as many keys as experts. This would assume that there could be as many sets of "optimal" decisions as there are experts. Each examinee's decision would be matched against each key to determine the key which produced the best match. This key could be used to score the examinee's decision. This model would assume either (1) no error in expert judgements or performance, or (2) that if there are errors in expert decisions then they are acceptable in the examinee. In both cases the main source of variation among experts is due to the nature of instruments and problems.

In the middle of the two extremes, it is possible to assume that variation in expert judgements and performances is due to both differences in experts' perceptions and decision-making skills, and to the different strategies of solving the problem (i.e., different sets of optimal decisions). In this model, it would be necessary to first group the experts into homogeneous subgroups and then to determine the "optimal" decisions within each subgroup. If the variation in scores is due to both differences in experts and to "optimal" decisions within the patient problem, the above model would require more than one key, but fewer keys than the number of experts.

In conclusion, current scoring procedures use a single key which may not accurately describe the consistent perceptions and/or selection of individual experts. On the opposite extreme of a single key, there may be as many keys as experts. All consistent perceptions among experts would be accurately described. However, since many experts will most likely share similar perceptions and/or selections, these experts may be divided into homogeneous groups and a key developed for each group. The extent to which identifying and not identifying homogeneous groups with similar perceptions and/or selections, has an effect upon inducing differences into examinee scores will be investigated in this study.

F. Scoring formulas

Different scoring formulas have been devised to summarize the pattern of examinee selections made on PMPs. Rumoldi (1955) uses an agreement and utility score. The agreement score is reported as the agreement between the optimal and chosen sequence of questions. The utility score is the average of weighted options selected by the examinee. Williamson (1965) developed five scores to summarize examinee selections: efficiency, proficiency, errors of omission, errors of commission, and a composite index of overall competence. The efficiency score is reported as the percentage of correct options selected to the total number of options selected. Proficiency is another name for the total test percentage score. Errors of omission is reported as the percentage of marks lost by failing to select correct options. Errors of commission is reported as the percentage of marks lost by selecting incorrect options. Failure to achieve 100 percent proficiency is by definition attributed to errors of omission and commission. A composite index of overall competence is reported as a weighted linear function of both efficiency and proficiency.

The National Board of Medical Examiners and the R.S. McLaughlin Examination and Research Centre use only one score (i.e., proficiency) to summarize examinee selections. Helfer et al. (1971), used five scores: process, diagnostic, efficiency, omission, and commission. The process score reflects the degree of match between examinee and expert's sequence of decisions. The diagnostic score reflects the accuracy of the diagnosis. The efficiency, omission and commission scores are similar to those devised by Williamson (1965). Freedman (1973) summarizes examinee selections by reporting the cost in dollars of hospitalizing, investigating, and treating a simulated patient. Berner et al. (1974), like the National Board and the R.S. McLaughlin Examination and Research Centre, summarize examinee selections by one score (i.e., proficiency).

Since it is important to thoroughly investigate the effects various scoring procedures have on summarizing examinee selections, four scores devised by Williamson (1965) will be used in this investigation. These four scores are: proficiency, error of omission, error of commission, and efficiency (see Appendix C for scoring formulas). The competence index for an unknown reason is not extensively used and will be excluded from this study.

3. Summary

A classification system was presented for categorizing PMP scoring techniques. Four scoring techniques currently used by two highly respected licensing agencies and one medical school were classified and compared. Similarities, differences, and possible short-comings of the four scoring techniques were outlined. Based on this background, eleven possible scoring procedures, four of which are currently used, are presented in the following chapter. The extent to which different scoring procedures induce differences in examinee scores will be investigated through the eleven scoring methods described in the following chapter.

CHAPTER IV

SCORING PROCEDURES

To devise and study various scoring procedures, the following characteristics were combined:

1. Classification

A. Categorizing options
  (1) group consensus
  (2) individual judgements
  (3) computer performance

B. Weights
  (1) constant
  (2) differential

C. Number of keys
  (1) single
  (2) multiple

Combining the above characteristics into various combinations produced eleven scoring procedures, four of which are currently used today. In order to describe the various scoring procedures, the following notation is employed.

2. Notation

A. Clinical Decisions in a Computer Simulation of a Patient Management Problem

In both linear and branching computer patient management problems, clinical decision points were represented as follows:

$$\text{CPMP K (DECISION)} = D_{11}, D_{21}, \ldots \ldots D_{pn} \quad (4.1)$$

where CPMP K (DECISION) represents all the decisions in the Kth CPMP, and

D_{pn} represents the ^{n}th decision point (or node) in the ^{P}th section.

B. Categorization of Clinical Decisions

To formulate a key, each clinical decision D_{pn} was categorized as either "definitely appropriate," "optional," or "definitely inappropriate" using group consensus, individual judgements, or computer performance. Clinical decisions classified as definitely appropriate were represented by a plus (+) sign, while those classified as definitely inappropriate were represented by a negative (-) sign. Clinical decisions classified as optional (i.e., neither (+) nor (-)) were represented as zero (0). Each clinical decision was further categorized according to its degree of appropriateness or inappropriateness in solving the patient's problem (i.e., +4, +3, +2, +1, 0, -1, -2, -3, -4). The number of plus or minus signs (i.e., +4 = ++++, +3 = +++, etc.) reflected the perceived degree of appropriateness or inappropriateness of the decision. Categorization of decisions by judges within a CPMP were represented as follows:

$$\text{CPMP K (CATEGORY)} = C_{111}, C_{112}, \ldots \ldots, C_{jpn} \quad (4.2)$$

where: CPMP K (CATEGORY) is the categorization (+, 0, -) of decisions in the Kth CPMP, and

C_{jpn} is the categorization of the jth judge in the pth section on the nth decision.

C. Weights of Clinical Decisions.

To construct a key, each categorized decision was weighted. The weights assigned tended to quantitatively reflect the degree to which each clinical decision was appropriate or inappropriate. These weightings (i.e., weight x categorization = weighting), when represented as a vector, were used as a key to calculate examinee scores. Weightings within a CPMP were represented as follows:

$$\text{CPMP K (WEIGHTING)} = W_{11}, W_{12}, \ldots \ldots, W_{pn} \quad (4.3)$$

where: CPMP K (WEIGHTING) is the weighting of decisions in the Kth CPMP, and

W_{pn} is the weighting assigned in the pth section to the nth decision.

D. Options Selected.

Examinee and expert selections of options within the Kth CPMP were represented as vectors of ones and zeros:

$$\text{CPMP K (SELECTION)} = S_{111}, S_{112}, \ldots \ldots, S_{ipn} \quad (4.4)$$

where CPMP K (SELECTION) is the selection made on the Kth CPMP, and

S_{ipn} is the selection made (*e.g.*, 1 = selected, and 0 = not selected) by the ith examinee or expert in the pth section on the nth decision.

Using the classification outlines in Section 1 and the notation provided in Section 2, the following eleven scoring procedures were investigated. For classifications of scoring procedures see Table 4.1 on page 64.

3. Eleven Scoring Procedures

A. Group consensus, constant weights, and single key (GCS).

This procedure for developing a key is currently being used by the National Board of Medical Examiners. Through group consensus each decision within a CPMP was categorized into one of three categories:

+ Category: it must be done for the well being of the patient,

0 Category: it is optional, (*i.e.*, a procedure that might or might not be done, depending upon local conditions and customs), and

- Category: it should definitely not be done, and, if done, would be a serious error in judgement that might be harmful to the patient

TABLE 4.1

Eleven Scoring Procedures Created by Varying 3 Methods of Categorizing Decisions, 2 Methods of Weighting and 2 Types of Keys

	Categorization of Options			Weights		Key	
Scoring Key	Group Consensus	Individual Judgements	Computer Performance	Constant	Different	Single	Multiple
GCS	X			X		X	
GDS	X				X	X	
ICS		X		X		X	
IDS		X			X	X	
ICM		X		X			X
IDM		X			X		X
CCS			X	X		X	
CDS			X		X	X	
CCM			X	X			X
CDM			X		X		X
Mc	X		X		X	X	

The categorizations by group consensus can be represented as follows:

$$\text{CPMP K (CATEGORY)} = C_{111}, C_{112}, \cdots \cdots, C_{gpn}$$

where C_{gpn} is the categorization (i.e., +, 0, or -) made by the gth group in the pth section on the nth decision.

To generate a scoring key each (+) and (-) categorization was assigned a constant weight of 1. The weightings assigned to each decision can be represented as follows:

$$\text{CPMP K (WEIGHTING)} = W_{111}, W_{112}, \cdots \cdots, W_{gpn}$$

where W_{gpn} is the weighting (i.e., +1, 0, or -1) assigned by the gth group in the pth section on the nth decision.

This vector of weightings became the scoring key.

B. Group consensus, differential weights, and single key (GDS).

This scoring procedure has been fostered by the College of Medicine at the University of Illinois and primarily used with branching PMPs. Through group consensus each decision was first categorized into one of five categories:

++ Category: choices which are CLEARLY INDICATED and IMPORTANT in the care of THIS patient at THIS stage in the workup or management,

+ Category: choices which are CLEARLY INDICATED but of a more ROUTINE nature (i.e., should be selected but are not of special significance in the care of THIS patient at THIS stage,)

0 Category: choices which are OPTIONAL, (i.e., the probability that they will be helpful for THIS patient at THIS stage is fairly remote or quite debatable,)

- Category: choices which are clearly NOT INDICATED though NOT HARMFUL in the management of THIS patient at THIS stage, and

-- Category: choices which are clearly CONTRA-INDICATED (i.e., are definitely harmful or carry an unjustifiable high cost in terms of risk, pain or money) in the care of THIS patient at THIS stage.

Clinical decisions categorized as either (++) = +2, or (--) = -2 were further categorized as either (+++) = +3, (++++) = +4, (---) = -3, or (----) = -4, depending upon their perceived degree of appropriateness or inappropriateness in solving the patient's problem.

A differential weight was assigned to each clinical decision according to its classification. For example, a clinical decision categorized as +3 was given a weight of $+(2^3) = +8$, and a clinical decision categorized as -4 was given a weight of $-(2^4) = -16$. The general formula used for deriving weights was $\pm(2^k)$ where k is the number of (+) or (-) signs. The vector of weightings became the scoring key.

C. Individual judgements, constant weights, and single key (ICS)

In this scoring procedure each expert independently categorized each decision into one of three categories as discussed in the GCS scoring key on page 63. There were as many categorizations as there were judges (J). To reduce the J categorizations to a single categorization that reflected the consensus of the group, decisions with relatively high interjudge agreement were identified and used to produce the key.

The following four steps were used to reduce the J categorizations and create a scoring key:

(1) count the number of times each decision was placed in the (+), (0), and (-) categories,

(2) select a criterion which reflects relatively high interjudge agreement (*i.e.*, .5J = 50% of judges),

(3) apply the criterion to the number of times each decision was placed in the (+) or (-) category. If the number equaled or exceeded the criterion, then the category was retained, otherwise, the decision was placed in the (0) category, and

(4) assign a constant weight of 1 to each (+) and (-) categorization.

D. Individual judgements, differential weights, single key (IDS).

In this scoring procedure each expert independently categorized and weighted decisions using the method explained under scoring key GDS on page 66. A single key was then produced by averaging the weights over judges. The key can be represented as follows:

$$\text{CPMP K (WEIGHTING)} = \frac{1}{J}\sum_{j=1}^{J}(W_{111}, W_{112}, \cdots \cdots, W_{jpn})$$
$$= \bar{W}_{11}, \bar{W}_{12}, \cdots \cdots, \bar{W}_{pn}$$

where CPMP K (WEIGHTING) is the derived key of weights for the Kth CPMP, and

$\bar{W}_{pn}$ is the average weight in the pth section, for the nth decision.

E. Individual judgements, constant weights, multiple key (ICM)

In this scoring procedure each expert independently categorized each decision into one of three categories and assigned a constant weight to the respective categories as discussed under scoring key GCS on pages 63 and 65. Using a centroid clustering procedure on weightings, the experts were divided into homogeneous groups. A scoring key was then produced for each subgroup which reflected relatively high interjudge agreement. The procedure explained under the ICS scoring key on page 67 (steps 1 through 4), was used to develop the scoring key for each subgroup.

The above procedure resulted in as many keys as there were homogeneous groups. If there were k homogeneous groups, then there were k keys. To score examinee selections, each of the k keys were used to calculate k sets of scores for each examinee. The key yielding the highest proficiency score was used to identify the subgroup which the examinee was most like.

F. Individual judgement, differential weights, multiple key (IDM).

In this scoring procedure each expert independently categorized and weighted each decision as discussed under scoring key GDS on pages 65 and 66. Like scoring key ICM a centroid clustering procedure was used to divide the experts into homogeneous subgroups according to differential weights assigned to options. To produce a single key for each subgroup, an average weight was calculated for each decision (see scoring key IDS on page 67 and 68). There were as many scoring keys as there were homogeneous groups. Given a total of k groups, each of k keys were used to calculate k proficiency scores for each examinee. The key yielding the highest score was used to identify the subgroup which the examinee was most like.

G. Computer performance, constant weights, single key (CCS).

This scoring procedure has been used by the R. S. McLaughlin Examination and Research Centre in their 1974 and 1975 paediatric examinations. Unlike the other scoring models, expert computer performance was used to categorize and weight clinical decisions. In this scoring procedure, experts, seeing the CPMP for the first time, took the examination. Their decisions were recorded and a scoring key developed based on the number of experts selecting each option. Expert selections are represented as follows:

$$\text{CPMP K (SELECTION)} = S_{111}, S_{112}, \ldots \ldots, S_{jpn}$$

where S_{jpn} is the selection made by the jth judge in the pth section on the nth option (*i.e.*, 1 = selected, and 0 = not selected).

To categorize options into three categories (*i.e.*, +, 0, or -) a criterion was chosen and applied to the proportion of experts selecting each option. Although any criterion may be chosen, that criterion used by the R. S. McLaughlin Examination and Research Centre was also used for this investigation.

Criterion:

$$+ \text{ Category: if } \frac{1}{J} \sum_{j=1}^{J} S_{jpn} \geq .5$$

$$0 \text{ Category: if } \frac{1}{J} \sum_{j=1}^{J} S_{jpn} < .5$$

$$- \text{ Category: if } \frac{1}{J} \sum_{j=1}^{J} S_{jpn} = 0$$

The categorized decisions were weighted as follows: the (+) category was assigned a weight of +1 and the (-) category was assigned a weight of -1.

H. Computer performance, differential weights, single key (CDS)

In this scoring procedure expert computer performance was again used to categorize options which were then differentially weighted. To categorize and weight options the selections for each option (*i.e.*, +1 = selected and -1 = not selected) were added over experts. This summation produced a number, ±J, for each option which was the weighting used in the scoring key.

I. Computer performance, constant weights, multiple key (CCM)

In this scoring procedure expert computer performance data were used to divide experts into homogeneous subgroups using a centroid clustering technique. Keys were developed for each subgroup by the method described under the CCS scoring procedure on page 70.

The categorized options within a subgroup were weighted ±1 and used as keys to calculate examinee scores. The key yielding the highest proficiency score was used to identify the subgroup which the examinee was most like and to calculate the final set of examinee scores.

J. Computer performance, differential weights, multiple key (CDM).

In this scoring procedure expert performance data were again used to divide experts into homogeneous subgroups using a centroid clustering technique. Categorizations and weights were developed for each group by the same procedure outlined under the CDS scoring key explained on page 80. The key yielding the highest proficiency score was used to identify the subgroup which the examinee was most like and to calculate the final set of examinee scores.

K. Group consensus and computer performance, differential weights and single key (Mc).

This scoring procedure has been used by the R. S. McLaughlin Examination and Research Centre. In this procedure a committee of experts collectively categorized the options into the three categories discussed under the GCS scoring key on page 63. Then expert problem-solvers who had not seen the CPMP before, took the examinations. Their selections were used to weight (+) and (-) options. Positive options were weighted by means of the following formula:

$$+W_{pn} = \frac{\sum_{j=1}^{J} S_{jpn}}{\sum_{j=1}^{J} \sum_{p=1}^{P} \sum_{n=1}^{N} S_{jpn}} \qquad (4.5)$$

where $+W_{pn}$ is the weight in the pth section on the nth (+) categorized option, and

S_{jpn} is the selection (*i.e.*, 1 = selected and 0 = not selected) made by the jth judge in the pth section on the nth decision.

The denominator of the above formula equals the total number of selections made by J experts on the (+) options. The numerator equals the number of judges who selected the nth decision in the pth section. Thus, the numerator divided by the denominator equals the proportion of selections made on the nth decision out of the total number of selections (*i.e.*, $\sum_{p=1}^{P} \sum_{n=1}^{N} W_{pn} = +1.0$).

The (-) decisions were weighted in a similar manner using the following formula:

$$-W_{pn} = \frac{J - \sum_{j=1}^{J} S_{jpn}}{JPN - \sum_{j=1}^{J} \sum_{p=1}^{P} \sum_{n=1}^{N} S_{jpn}} \qquad (4.6)$$

where W_{pn} is the weight for the nth categorized option in the pth section.

The denominator of the above formula equals the total number of selections NOT made on (-) options. The numerator equals the number of judges who did NOT select the nth decision in the pth section. Thus $-W_{pn}$ is the proportion of selections NOT made on the nth decision out of the total number.

In summary, a scoring key was generated by categorizing options using the group consensus method and assigning weights by computer performance

L. Author, differential weights, single key (author).

A twelfth scoring key was also investigated. These were the scoring keys generated by the authors of the CPMPs and used to calculate final examinee results. The author's key had the following classification:

i) method of categorization: author
ii) weight : differential
iii) key : single

This scoring key will be referred to as 'author'.

Table 4.2 summarizes the twelve scoring keys that were studied in this investigation.

Table 4.2

Categorization of Proposed Scoring Procedures

Acronym	Categorization of Options	Weights	Key
Author	author	differential	single
GCS	group consensus	constant	single
GDS	group consensus	differential	single
ICS	individual judgement	constant	single
IDS	individual judgement	differential	single
ICM	individual judgement	constant	multiple
IDM	individual judgement	differential	mulitple
CCS	computer performance	constant	single
CDS	computer performance	differential	single
CCM	computer performance	constant	multiple
CDM	computer performance	differential	multiple
Mc	computer performance group consensus	differential	single

4. Assumptions and Limitations Underlying Scoring Procedures

The above scoring keys were based upon various assumptions. These assumptions and their limitations will be discussed according to method of categorization, weight and key.

A. Method of Categorization

a. Group consensus (i.e., GCS and GDS)

This method assumed that:

1) variations in categorization were due to differences among experts, which could "best" be resolved through group discussion, and

2) knowing the correct solution to the problem resulted in the "best" categorization of options.

The above assumptions are subject to the following limitations:

1) the ability to resolve differences among experts was dependent upon the dynamics of the group and was limited by the extent to which the group was able to collectively work together, and

2) a scoring key produced by categorizing options while knowing the correct solution may not model problem-solving behavior

b. Individual judgements (i.e., ICS, IDS, ICM and IDM).

This method assumed that variations in categorizations were due to differences among experts which could "best" be resolved by either categorizing options using "high" interjudge agreement or by averaging weightings assigned to each option over judges.

The above assumption may be limited by the extent to which individual experts share the views and judgements of other experts. In addition, this method, like group consensus, may be limited by the differences in tasks performed by expert judges and problem-solvers.

c. Computer performance (i.e., CCS, CDS, CCM and CDM).

This method assumed that:

1) variations in categorization could "best" be resolved by either categorizing options using "high" interjudge agreement or by summing selections over judges (i.e., +1 = selected, -1 = not selected), and

2) examinee scores would more closely reflect the decisions of expert problem-solvers rather than that of the expert judges.

Assumption 1 above may be limited the extent to which individual experts share the views and opinions of other experts.

B. Weights

a. Constant (i.e., GCS, ICS, ICM, CCS and CCM).

When employing constant weights it was assumed that all (+) and (-) decisions were equally appropriate or inappropriate in solving the patient's problem.

Scores generated under the above assumption may be limited by the extent to which this is indeed true. For example, given a hypothetical CPMP with three decisions:

CPMP K (DECISION) = D_{11}, D_{12}, D_{13}

CPMP K (CATEGORY) = +   -   +

CPMP K (WEIGHTING)= +1   -1   +1 = KEY

and the selections of three hypothetical examinees:

CPMP K $(\text{SELECTION})_1$= 1   1   1

CPMP K $(\text{SELECTION})_2$= 0   0   1

CPMP K $(\text{SELECTION})_3$= 1   0   0

the above key would lead to identical examinee scores in spite of the different response patterns.

b. Differential (i.e., Author, GDS, IDS, IDM, CDS, CDM and Mc).

When employing differential weights it was assumed that all (+) and (-) decisions were not equally appropriate or inappropriate for solving the patient's problem.

The above assumption may be limiting

to the extent by which differential weights do or do not reflect the appropriateness/inappropriateness of decisions.

C. Key

a. Single (i.e., Author, GCS, GDS, ICS, IDS CCS, CDS and Mc).

The single key assumed that there was:

1) only one set of "correct" clinical decisions and,

2) only one "optimal" route through the patient management problem.

The use of single keys may be straightforward but is limited to the extent that it ignores the possible consistent, but different, perceptions among individual experts.

b. Multiple (i.e., ICM, IDM, CCM and CDM).

The multiple key assumed that:

1) there were consistent, but different, perceptions among experts which could be isolated using the centroid clustering technique, and

2) the scoring key producing the highest proficiency score could be used to identify the subgroup which the examinee was most like.

This procedure is limited to the extent to which it is possible to subdivide a group into homogeneous clusters.

5. Summary

Twelve scoring procedures, four of which are currently in use, have been outlined. The underlying assumptions on which the scoring procedures are built have been identified along with the possible limitations due to the inappropriateness of assumptions made. The following chapter presents the materials and methods used in this investigation.

CHAPTER V

MATERIALS AND METHODS

1. Subjects

Data for this investigation was gathered by permission and cooperation of the R. S. McLaughlin Examination and Research Center. In May of 1976, 111 medical students, who had completed four years of medical training at the University of Alberta, wrote four CPMPs as part of their certifying examinations. These examinations were administered using the IBM 1500 computing facilities operated by the Division of Educational Research Services, Faculty of Education, University of Alberta

2. Examinations (CPMPs)

Two linear and two branching CPMPs were selected for investigation in this study: CPMP 1 (linear) represented a 44 year old man with a cardiac problem; CPMP 2 (linear) simulated a 56 year old man with anemia of unknown origin; CPMP 3 (branching) involved a 25 year old female with a gynecological problem, and CPMP 4 (branching) simulated a 21 year old female with an obstetrical problem.

A. CPMP 1.

CPMP 1 tested the candidate's ability to manage a patient with a heart problem. The problem was broken down into nine sections with questions under each section as indicated in Figure 5.1 (i.e., section 1-1, questions lettered a-i). The nine sections were presented as follows:

Section 1-1: initial presenting problem - what is an appropriate hypotheses?

Section 1-2: patient admitted to hospital - what laboratory investigations should be undertaken?

Section 1-3: based on the laboratory results obtained - what management should be undertaken?

Section 1-4: patient's condition becomes critical - what management should be undertaken?

Section 1-5: patient's condition improves, electrocardiogram(ECG) presented - which arrhythmia is presented?

Section 1-6: given interpretation of ECG - what management should be undertaken?

Section 1-7: patient's original problem corrected but now could be developing a new problem (i.e., possible left ventricular failure) - on physical examination what would be the expected results?

Section 1-8: management of left ventricular failure?

Flowchart of CPMP 1
A 44 year old man with a cardiac problem

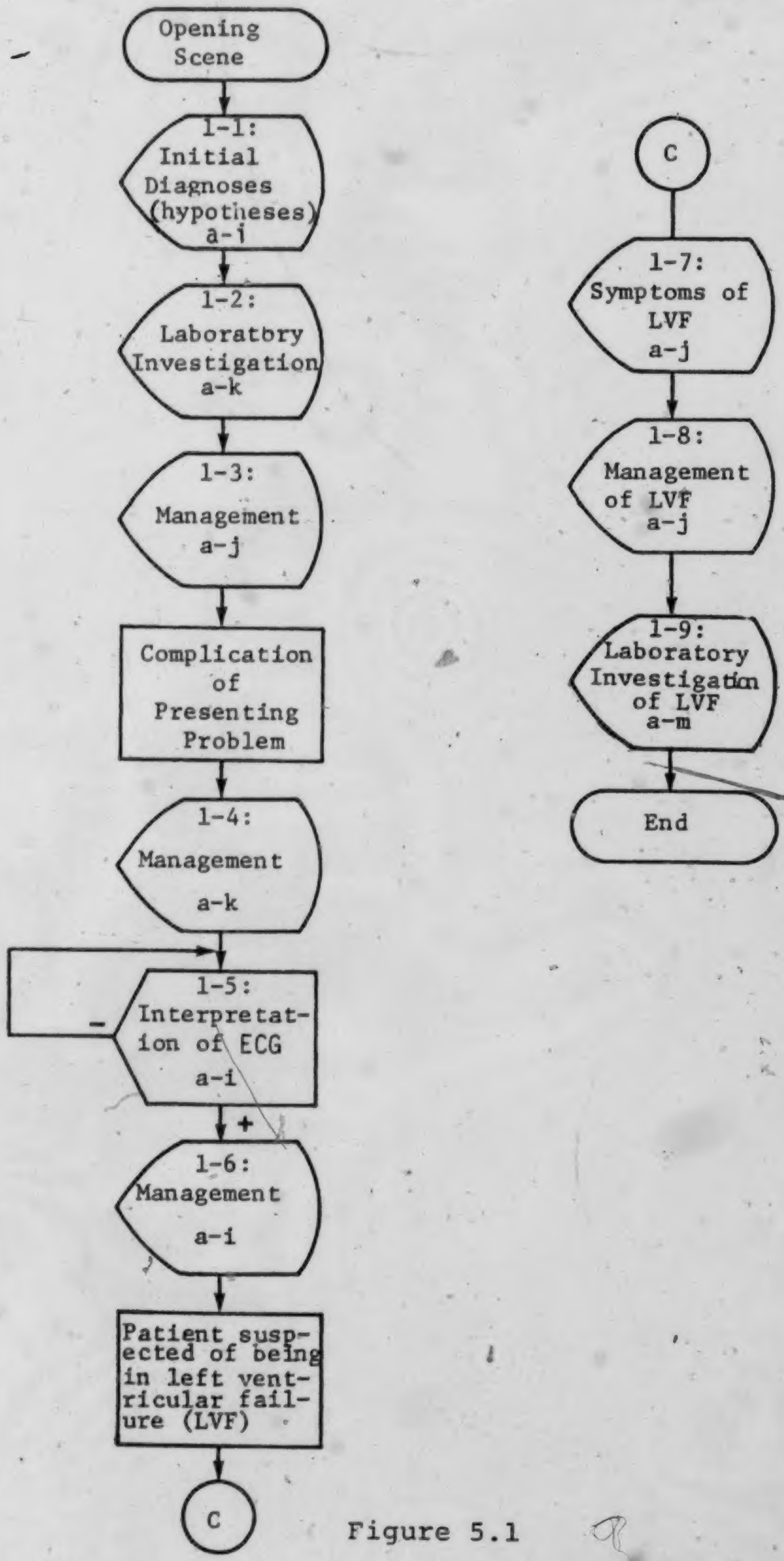

Opening Scene
1-1: Initial Diagnoses (hypotheses) a-i
1-2: Laboratory Investigation a-k
1-3: Management a-j
Complication of Presenting Problem
1-4: Management a-k
1-5: Interpretation of ECG a-i
−
+
1-6: Management a-i
Patient suspected of being in left ventricular failure (LVF)
C
C
1-7: Symptoms of LVF a-j
1-8: Management of LVF a-j
1-9: Laboratory Investigation of LVF a-m
End

Figure 5.1

Section 1-9: patient recovers - what investigative procedures should be undertaken?

The structure of CPMP 1 is linear (i.e., the candidate proceeded from one section to the next until the problem was completed). Only in Section 1-8 was the candidate's response directly dependent upon the information obtained in the preceding section.

B. CPMP 2.

CPMP 2 tested the candidate's ability to manage a middleclass, 56 year old male with anemia of unknown origin. The problem was broken down into the following nine sections and sub-sections (see Figure 5.2):

Section A1: initial presenting symptoms - based upon results of initial laboratory and physical examination - what further investigations should be ordered?

Section A2: upon establishing correct tentative diagnosis - what investigative procedure should be undertaken?

Section A3A: if candidate administers three units whole blood, patient develops pulmonary edema - what corrective measures should be undertaken?

Section A4: candidate still without diagnosis - what additional investigative procedures should be ordered?

Section A5: based upon findings of A4 - what is appropriate choice of treatments?

Flowchart of CPMP 2

A 56 year old man with Anemia of unknown origin

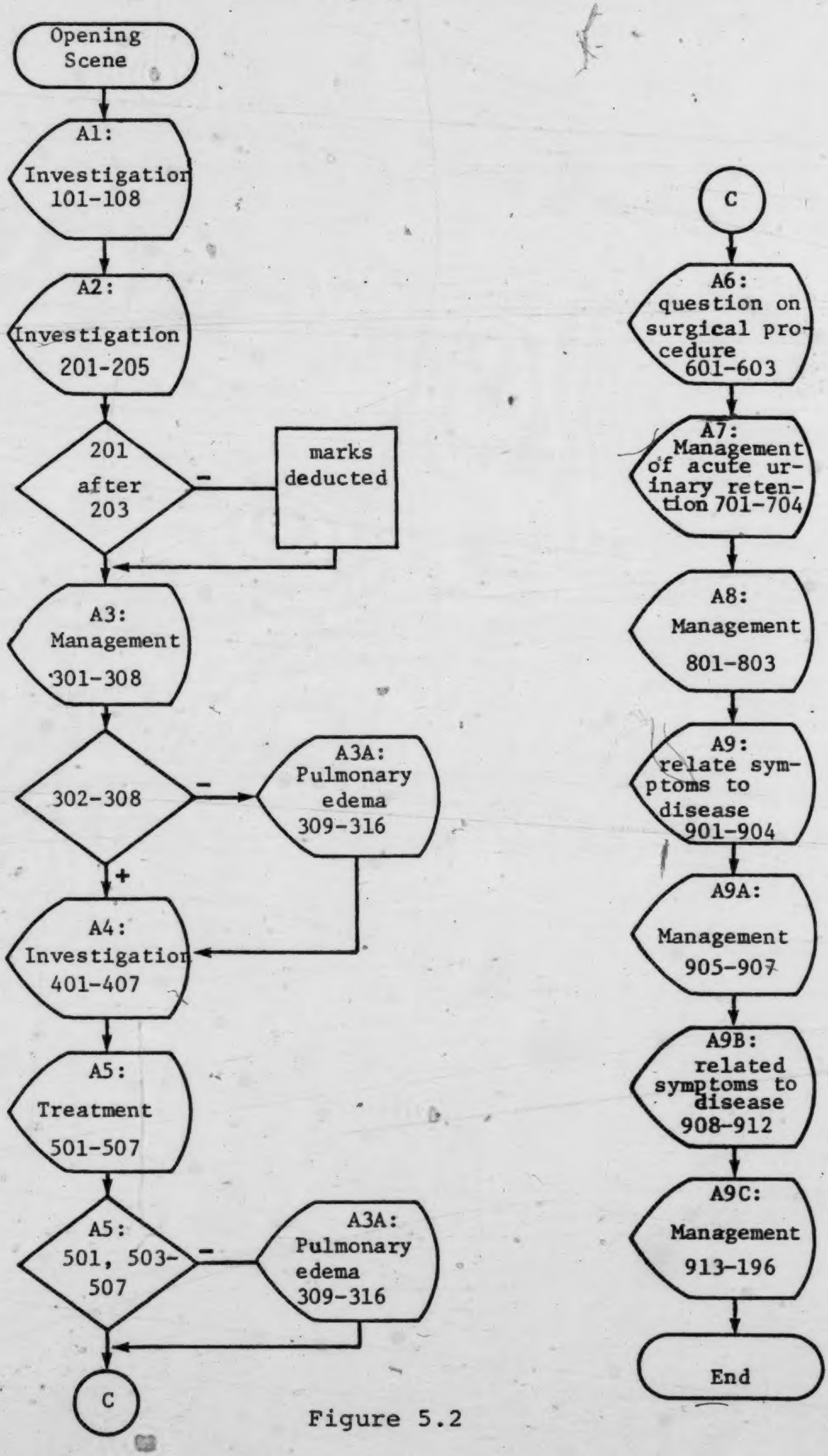

Opening Scene
A1: Investigation 101-108
A2: Investigation 201-205
201 after 203
-
marks deducted
A3: Management 301-308
302-308
-
+
A3A: Pulmonary edema 309-316
A4: Investigation 401-407
A5: Treatment 501-507
A5: 501, 503-507
-
A3A: Pulmonary edema 309-316
C
C
A6: question on surgical procedure 601-603
A7: Management of acute urinary retention 701-704
A8: Management 801-803
A9: relate symptoms to disease 901-904
A9A: Management 905-907
A9B: related symptoms to disease 908-912
A9C: Management 913-196
End

Figure 5.2

Section A3A: if candidate administers three units whole blood, patient develops pulmonary edema - what corrective measures should be undertaken?

Section A6: surgery required - what is correct surgical procedure?

Section A7: acute urinary retention develops - what management should be undertaken?

Section A8: two weeks pass and patient still unable to void - what management should be undertaken?

Section A9: two of patients' problems corrected - what management for remaining problems?

Section A9A: after correct diagnosis of two of the remaining problems - what directions to patient?

Section A9B: what is most likely cause of remaining problem?

Section A9C: given laboratory results of remaining problem - what directions to patient?

The above structure of CPMP 2 is presented as a flowchart in Figure 5.2. The structure is primarily linear. Only in Sections A3 and A5 could the candidate's linear flow be disrupted to Section A3A for corrective management.

C. CPMP 3.

CPMP 3, the gynecological problem, from a branching point of view, was the most complex of the four simulations

FLOWCHART OF CPMP3: A 25 year old female with a gynecological problem.

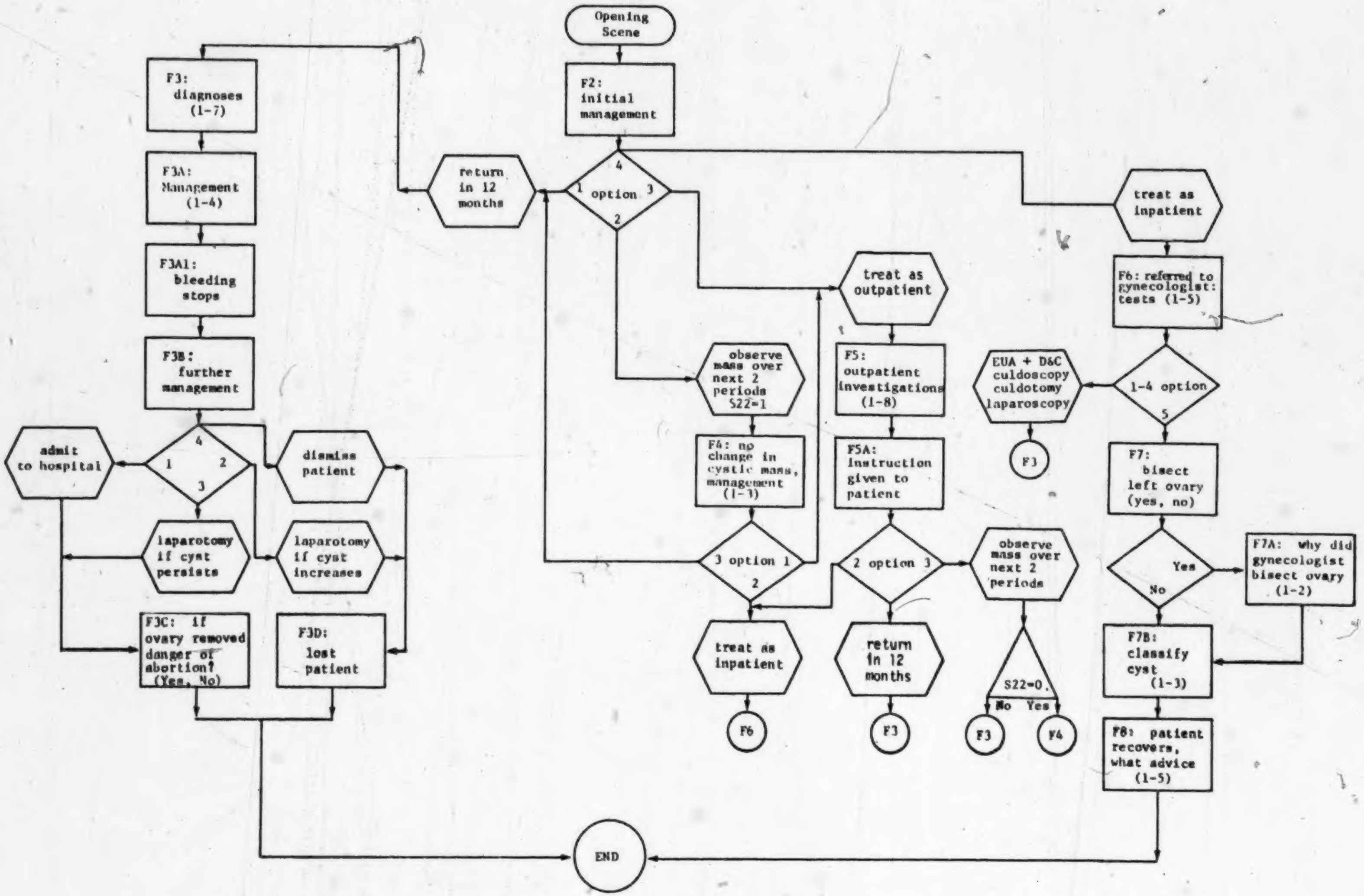

Opening Scene
F2: initial management
4
1 option 3
2
return in 12 months
treat as inpatient
treat as outpatient
observe mass over next 2 periods S22=1
F4: no change in cystic mass, management (1-3)
3 option 1
2
treat as inpatient
F6
F5: outpatient investigations (1-8)
F5A: instruction given to patient
2 option 3
return in 12 months
F3
observe mass over next 2 periods
S22=0.
No
Yes
F3
F4
F6: referred to gynecologist: tests (1-5)
1-4 option
5
EUA + D&C culdoscopy culdotomy laparoscopy
F3
F7: bisect left ovary (yes, no)
Yes
No
F7A: why did gynecologist bisect ovary (1-2)
F7B: classify cyst (1-3)
F8: patient recovers, what advice (1-5)
F3: diagnoses (1-7)
F3A: Management (1-4)
F3A1: bleeding stops
F3B: further management
4
1
2
3
admit to hospital
dismiss patient
laparotomy if cyst persists
laparotomy if cyst increases
F3C: if ovary removed danger of abortion? (Yes, No)
F3D: lost patient
END

Figure 5.3

(see Figure 5.3). There is a left and right side to Figure 5.3 with several connecting pathways between. The left side of the figure could be called the mismanagement side and the right, the correct management side.

CPMP 3 tested the candidate's ability to manage a female patient with a cystic mass in the region of the ovary. The candidate's perception of the significance of the cyst was tested on three occasions (i.e., F2, F4 and F5a). If the candidate perceived the mass as nonsignificant and treated the situation as an out-patient problem, the candidate was branched to the mismanagement side. If, on the other hand, the candidate perceived the cyst to be significant and hospitalized the patient or treated the situation as an in-patient problem, the candidate was branched to the correct management side. One additional test was given to these candidates to determine whether they remained on the correct management side. This test occured in Section F6 where the candidate was told that the patient was admitted to hospital and referred to a gynecologist; the candidate was then asked what investigations the gynecologist would be expected to undertake. If the candidate did not choose laparotomy, he or she was branched to the mismanagement side. On the other hand, selecting the laparotomy resulted in the candidate staying on the correct management side. Therefore, in order to remain on the correct management side, the candidate had to recognize the cyst as being significant, hospitalize the patient, and know that a laporotomy should be performed by the gynecologist.

D. CPMP 4.

CPMP 4, the second of the branching problems, tested the candidate's ability to handle a prolonged delivery problem. The first section tests the candidate's ability to select the appropriate investigative procedures (i.e., Section F2). The remaining six sections deal with the management of the patient's problem (i.e., Sections F3, F6, F8, F11, F16, and F19). Within the management section, if a candidate chose to refer the patient to an obstetrician, the clinical encounter ended. For a diagramatical representation of the structure of the obstetrical problem see Figure 5.4.

E. Comparison of CPMPs.

The four CPMPs differ both in content and structure. Firstly, CPMP 1 represented a cardiac problem; CPMP 2, an anemia problem; CPMP 3, a gynecological problem; and CPMP 4, an obstetrical problem.

Secondly, the CPMPs differed in the stage of intervention and the urgency of treatment. The obstetrical and gynecological problems required immediate intervention while the anemia and cardiac problems did not.

A FLOWCHART OF CPMP4: A 24 year old female with an obstetrical problem

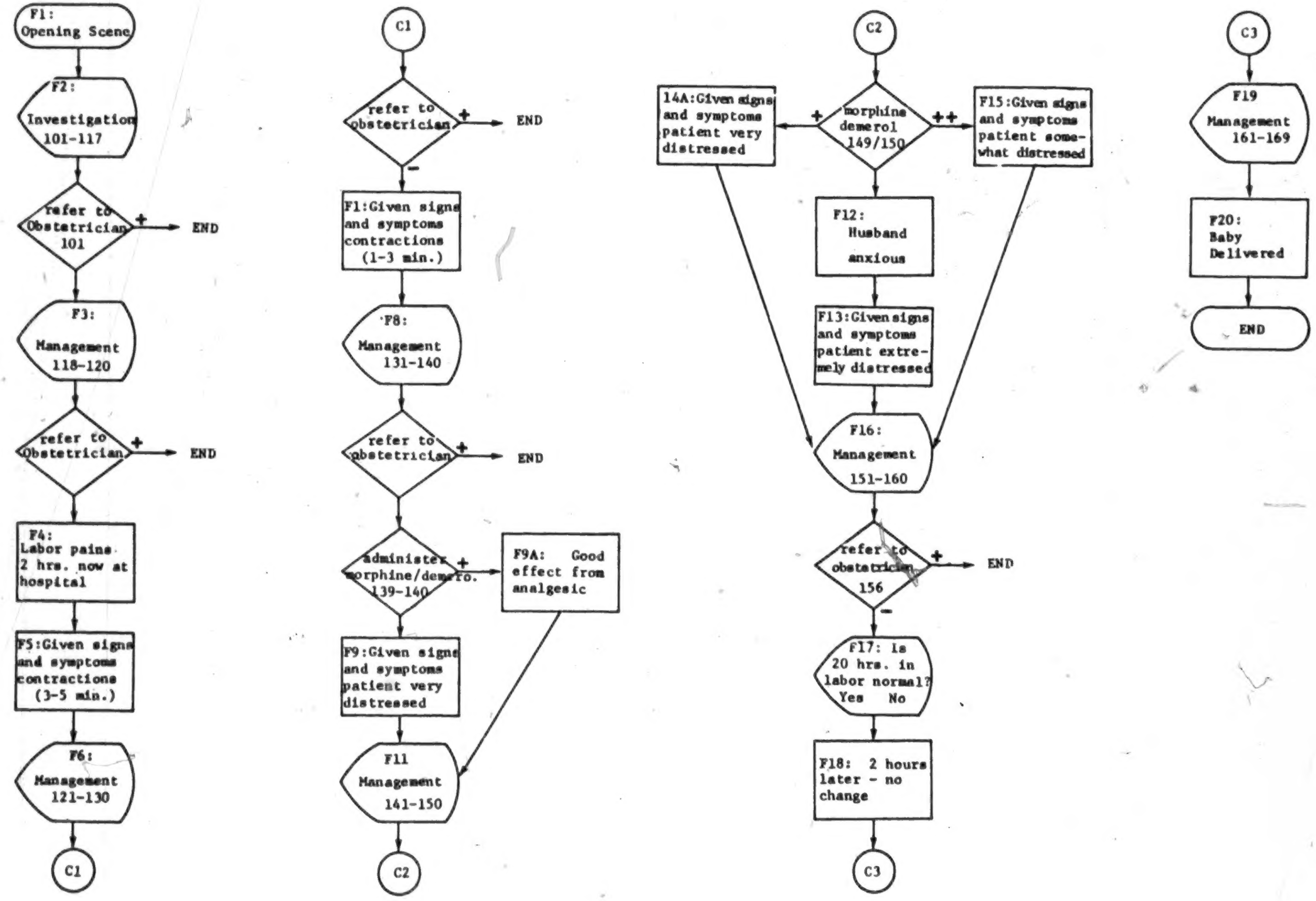

F1: Opening Scene
F2: Investigation 101-117
refer to Obstetrician 101
+
END
F3: Management 118-120
refer to Obstetrician
+
END
F4: Labor pains 2 hrs. now at hospital
F5: Given signs and symptoms contractions (3-5 min.)
F6: Management 121-130
C1
C1
refer to obstetrician
+
END
-
F1: Given signs and symptoms contractions (1-3 min.)
F8: Management 131-140
refer to obstetrician
+
END
administer morphine/demero. 139-140
+
F9A: Good effect from analgesic
F9: Given signs and symptoms patient very distressed
F11 Management 141-150
C2
C2
morphine demerol 149/150
+
14A: Given signs and symptoms patient very distressed
++
F15: Given signs and symptoms patient somewhat distressed
F12: Husband anxious
F13: Given signs and symptoms patient extremely distressed
F16: Management 151-160
refer to obstetrician 156
+
END
-
F17: Is 20 hrs. in labor normal? Yes No
F18: 2 hours later - no change
C3
C3
F19 Management 161-169
F20: Baby Delivered
END

Figure 5.4

Thirdly, the CPMPs differed in the interventions offered. For example, CPMP 1 had one section on hypothesis, two sections on laboratory investigations, three sections on treatment, one section on correct interpretation of ECG and one section on physical examination. CPMP 2 had three sections on investigations, seven sections on treatment, and two sections relating presenting signs and symptoms to the most likely diagnosis. CPMP 3 had three sections which tested the candidate's perception of the significance of the ovarian cyst, two sections on investigations, one on the correct investigative procedure of the gynecologist, one on whether the gynecologist would bisect the left ovary, one on the type of cyst, and one on advice to be given to the patient. CPMP 4 had one section on investigative procedures and six on treatment.

The CPMPs also differed in the type of feedback candidates received. Some feedback was corrective (_i_._e_., CPMP 4, option #129, "administer morphine, 10 mg.", given answer, "not indicated"), other feedback was only confirmatory (_i_._e_., CPMP 4, option #123, "take blood for cross-match", answer, "done"). In CPMP 2, candidates were allowed to answer some sections until the correct answer was found. Sometimes corrective feedback for previous sections was given at the beginning of the next section.

Lastly, CPMPs differed in the number of questions with single and multiple answers. In CPMP 1, there were nine multiple response items; in CPMP 2, twelve (four were actually single response items but the candidate was allowed to respond until the correct answer was found); in CPMP 3, there were six single and four multiple response items; and in CPMP 4, there were two single and six multiple response items.

There were few features common to all CPMPs, those features being that all CPMPs required candidates to make selections from a list of options and all were administered by computer.

3. "Expert" Physicians

The Edmonton area was canvassed for physicians who would volunteer to take part in this study. Those who volunteered were offered a choice of participating on one of three days. Physicians participating on day one constituted Group A; on day two, Group B; and, on day three, Group C. Although volunteers were not randomly assigned to each of the three groups, every effort was made to make the three groups as homogeneous as possible. For purposes of this study, it is assumed that the three groups were equal in medical training, education, years of practice and age. The composition of the groups is summarized in Table 5.1 below:

TABLE 5.1

Biographical Data of
"Expert" Physicians

		Group		
		A	B	C
1.	Number of participants	10	16	11
2.	Number of males	9	13	9
3.	Number of residents	1	3	1
4.	Number of practicing specialists	3	3	1
5.	Number of practicing family practicioners	6	10	9
6.	Average number of years in practice	13.3	10.1	7.2
7.	Average age	38.4	38.9	37.1

A total of 37 "expert" physicians took part in this study. The large number of physicians participating can only be attributed to the concerted efforts made by the staff at the R. S. McLaughlin Examination and Research Centre.[1]

4. The Development of Scoring Keys

To develop the scoring keys, the physicians were asked to categorize options using the three methods of categorization explained in Chapter IV, namely, computer performance, group consensus, and individual judgement. Weights were then assigned to options on a constant or differential basis.

---

[1] Special thanks is extended to Wayne Osbaldeston of the R. S. McLaughlin Center who played a key role in the data gathered.

Since many of the physicians who took part in the study had never been exposed to a computer patient management problem (CPMP), it was firstly necessary to demonstrate a CPMP and explain the basic computer terminal procedures required to interact with the system. The physicians were then asked to sign-on to a given CPMP and select all options that would be helpful in resolving the simulated patient's problems. Selection of options through direct interaction with the computer was referred to as the 'computer performance' method of categorization of options.

Next, the physicians were given a short but thorough course on categorizing and weighting options (see instruction sheets for linear problems entitled Appendix D, and branching problems, entitled Appendix E). The physicians were then given a different CPMP (i.e.,CPMP 1, 2, 3 or 4) and asked to categorize the options using the group consensus method of categorization. Upon completion of this task, the physicians were given another CPMP and asked to independently categorize options (i.e., the individual judgement method of categorization). The above order of activities was given to Groups A, B, and C respectively on day one, two and three.

The activities allocated to the specific groups are illustrated in Table 5.2.

TABLE 5.2

Tasks and CPMPs Performed by the Three Groups of "Expert" Physicians

	GROUP		
	A	B	C
CPMP 1	Consensus	Judgement	Performance
CPMP 2	Performance	Consensus	Judgement
CPMP 3	Performance	Judgement	Consensus
CPMP 4	Judgement	Performance	Consensus

The data gathered from the expert physicians in the above course of activities were used directly to construct seven of the eleven (excluding the Author scoring key) scoring keys employed in this study, namely, the GDS IDS, IDM, CCS, CDS, CCM and CDM keys (see page 79). However, the remaining four scoring keys, namely, the GCS, ICS, ICM and Mc keys, could not be directly developed using the above data. These keys required each option to be placed into one of three categories (i.e., +, 0, and -). The above seven keys, however, had options which had been assigned nine weightings (i.e., weightings = categorization + weight), these weightings being +16 +8. +4, +2, 0, -2, -4, -8, and -16. In order to establish keys for the GCS, ICS, ICM and Mc keys, the nine weightings had to be reduced to three. To carry out this reduction, the following reduction rules were applied:

+ Category:	all options categorized as positive (i.e., +2, +4, +8, and +16 were placed in the (+) category,
0 Category:	all options categorized as zero remained zero, and
- Category:	all options categorized as negative (i.e., -2, -4, -8 and -16 were placed in the (-) category.

In order to carry out this reduction it was assumed that there would be an insignificant difference between reducing categories and having experts classify options using only three categories.

5. Description of Data

The twelve scoring keys were firstly re-scaled so that the maximum true/false proficiency score equaled 100%. Each of the re-scaled keys was then used to calculate the following four performance scores for each CPMP: proficiency, error of commission, error of omission, and efficiency. (See Appendix B).

As a result, for each examinee, 192 scores were calculated (*i*.*e*., 4 scores X 4 CPMPs X 12 scoring procedures).

6. Method of Data Analysis

A. Reliability Measures.

Classical reliability theory is based upon the assumption that every test has a true score; belongs to only one family of parallel tests (i.e., items are homogeneous); and, is unique depending on the partitioning of variance. However, the nature of the patient management problems made it necessary to consider a variety of aspects of reliability. In the CPMPs used in this study:

1) examinees could be directed to skip entire sections either because they successfully avoided complications, or because they took a different pathway in solving the patient problem,

2) the selection of an item could provide information about the problem not available to others who had not selected that option, and

3) the number of options selected may be more a reflection of the personality of the examinee than of the correctness or incorrectness in arriving at the solution to the problem. Thus, the very structure of the simulations, their content and use, suggested that reliability be treated as a multi-dimensional concept. Cronbach (1963)

treated reliability as that attribute of measurement which is related to "generalizability" of response. Cattell (1964) advocated that the "consistency of measurement be used as a concept to replace the more vague term of "reliability". According to Cattell, (1964), the "consistency of measurement" has at least three aspects:

1) consistency across occasion,
2) consistency across tests, and
3) consistency across people.

Consistency across occasion refers to:

1) the degree of agreement in results obtained from different scoring procedures,

2) the property usually referred to as test-retest reliability, and

3) the difference in results produced by different conditions of administration.

Consistency across tests refers to agreement in the results of two tests that purport to measure the same trait.

Finally, consistency across people refers to the appropriateness of a test in measuring the same trait in samples.

This study focused on consistency across occasions being:

1) the degree of agreement in results obtained from different scoring procedures,

2) the test-retest reliability of homogeneous parts within each CPMP.

In addition, the consistency of data used in constructing the scoring keys was assessed by analyzing the selections of expert problem-solvers and judgements of expert raters.

a) Consistency across occasion.

The consistency across occasions was assessed by determining whether different properties of the scoring key altered the:

1) distribution of scores (i.e., skewness, kurtosis, and variance).,

2) linear relationship among scores,

3) mean scores

4) absolute level of examinee performance

5) test-retest properties of homogeneous items within tests, and

6) rank order of examinees.

The above six variables were analyzed as follows:

1) distribution of scores: changes in skewness and kurtosis were assessed by comparison over scoring procedures; no statistical test was used. The F statistic was used to compare the variance of scores among scoring procedures;

2) linear relationships among scores: a principal components factor analysis with iterations was used to determine the underlying components of the CPMP scores; these components were rotated by the varimax technique and the factor loadings used to interpret the linear relationship among scores;

3) mean scores: a one-way multivariate analysis with repeated measures over scoring procedures and CPMPs was used to determine the effect that scoring procedures had upon mean scores;

4) absolute level of performance: examinee proficiency scores generated by the twelve scoring procedures, were compared against a minimal level of performance (MPL). The changes in satisfactory (pass)/unsatisfactory (fail) status over scoring procedures was examined. The method of arriving at the MPL is elaborated on page 101;

5) the test-retest properties of homogeneous items within CPMPs: a Cronbach's alpha and Lord's maximum alpha were used to estimate test-retest reliability. The method of reducing the CPMPs to homogeneous items is discussed on page 102;

6) rank order of examinee: a z statistic for dependent samples was used to determine variation among examinee rankings which were induced by the scoring procedures. An attempt was made to link these observed changes to the properties of the scoring procedures.

b. Estimation of a Minimal Standard of Performance (MPL)

Although there is a great deal of controversy surrounding the utility and the methods of determining a criterion level of performance, the practice is advocated by the Centre for the Study of Medical Education at the Univeristy of Illinois and has been adopted by medical schools on the pass/fail system of grading, one such medical school being the University of Calgary. For this reason, it was felt that it was important to investigate the effect that various scoring procedures could have upon altering examinee satisfactory (pass)/unsatisfactory (fail) status.

The method selected to calculate the MPL was devised at the University of Illinois:

MPL = 100 X (Sum of "indispensable positives") – (Sum of "forgiveable" negatives)/ Maximum score[2]

This method was designed for use with differential weights and applied to the following scoring procedures: Author, GDS, IDS CDS, and Mc. Since the MPL was to be applied to the scores generated by the twelve scoring procedures, it was felt that the MPL should not reflect the decision of any one procedure. Therefore, one MPL was calculated for each CPMP. This was achieved by calculating an MPL for the five procedures, averaging the five MPLs for each CPMP, and rounding the MPL off to the nearest 5%. This number was assumed to reflect the absolute minimum standard of performance for the test.

---

[2] The total number of marks that could be accumulated by optimal choices.

c. Estimating the Reliability of Homogeneous Options Within CPMPs.

In order to estimate the test/retest reliability of options within tests, each CPMP was firstly reduced to its homogeneous items. This was achieved by dividing options into two groups:

1) history-taking, laboratory investigation and physical examination, and

2) management or treatment.

This type of grouping was supported by the findings of Donnelly et al (1974), Juul et al (1977) and Skakun (1978), who concluded that these skills (i.e., data gathering and management skills) underlie the solution of clinical problems.

Secondly, the grouping with the largest number of options was selected within each CPMP.

Thirdly, two estimates of the reliability coefficient were calculated on examinee responses to these homogeneous options:

1) Cronbach's coefficient alpha, and

2) Lord's formula for maximizing the coefficient alpha.

Cronbach's coefficient seemed appropriate for estimating the degree of generalizability of the data gathering or management skill from one test to tests containing the same clinical problem. Lord's formula was employed to obtain a maximum limit for the estimated parameter.

B. Validity Measures

Since the CPMPs, having been obtained from the R. S. McLaughlin Examination and Research Centre, had been field tested and were administered as part of the fourth year final examinations in the Faculty of Medicine at the University of Alberta, it was assumed that the CPMPs possessed content, construct and concurrent validity. To determine the validity of the expert problem-solvers' selections, their scores were compared to the MPL. For the expert's scores to be valid, it was expected that they would be higher than those of the examinees' and that no expert would score below the minimum pass level (MPL).

The results of this study are presented in the following chapter.

CHAPTER VI

RESULTS

1. Characteristics of Scoring Keys

As discussed in preceding chapters, the twelve scoring procedures were made up of combinations of methods of categorization, weights and single and multiple keys. The method of categorization determined which options were categorized as positive, neutral and negative. The weight assigned to these options plus the categorization determined the weighting assigned to the option within the scoring key, (categorization x weight = weighting). It was observed that the above scoring keys produced different weights for the same option and, therefore, different scoring keys for the same CPMP. For example, in CPMP 1, the author procedure weighting of option #1-1-d, "acute anxiety state", was +0.8% while the computer procedure weighting of the same was -1.0%. It was also observed that the same options were given large weights but categorized as opposites. For example, the author method categorized option #1-4-j of CPMP 1, "give Heparin 5000 units by intervenous infusion q4h", as a definite course of action to be taken and gave it a weighting of +1.6%. However, the computer method categorized it as an action to be definitely avoided and gave it a weighting of -1.6%. It follows from the above that different weightings for options among scoring keys could result in different examinee scores for the same CPMP.

The extent to which scoring key categorizations and weights differed would in turn effect the extent to which scores differed. Differences in categorizations and weights are presented in Tables 6.1 and 6.2. Table 6.1 presents the number of negative and positive options by CPMP and scoring key. From the table it is evident that the number of positive and negative options varied greatly within the same CPMP. For example, in CPMP 1, the number of negative options decreased from 45 in the GCS scoring key to 20 in the CCS scoring key. The extent to which the weights differed between scoring keys is observed in Table 6.2. Table 6.2 presents the percentage of weights attributed to negative options. Once again, a large variation is observable. For example, in CPMP 1, 66.5% of the weight was assigned to negative options in the CDS scoring key but only 36.6% in the CCS key.

It is also of interest to note that there was a relationship between the structure of the CPMP and the average

Table 6.1

Number of Negative/Positive Options

By CPMP and Scoring Key

CPMP	Number of Negative/Positive Alternatives												TOTAL
	Author	GCS	GDS	ICS	IDS	ICM	IDM	CCS	CDS	CCM	CDM	Mc	
1	47*/39**	45/43	45/43	36/41	47/44	27/30	50/41	20/37	56/36	29/33	54/33	45/43	92
2	40/36	42/33	42/33	18/37	30/46	19/38	26/51	26/26	51/24	28/26	51/22	42/33	77
3	1/20	9/13	9/13	4/14	7/17	2/14	7/18	7/12	7/15	8/14 7/9	9/13 6/8	9/13	32
4	20/29	19/30	19/30	16/28	19/28	14/29	18/34	13/29	34/18	13/20	37/15	19/30	52

*number of negative options
**number of positive options

Table 6.2

Percentage Weight Out of a Scale of 100 Percent

Related to Negative Options

CPMP	Author	GCS	GDS	ICS	IDS	ICM	IDM	CCS	CDS	CCM	CDM	Mc	Average
1	53.0	51.0	49.5	47.0	42.0	46.5	44.0	36.5	66.5	48.5	67.0	50.0	50.0
2	55.5	59.0	52.0	35.5	33.0	36.0	32.0	45.5	67.0	45.0	67.5	50.0	48.2
3	20.0	41.0	41.5	22.0	28.0	13.5	26.5	37.0	51.5	23.5 50.0	21.5 48.5	50.0	33.9
4	55.0	39.0	41.5	36.4	40.5	32.5	39.5	31.0	67.5	44.5	65.0	50.0	44.6
Average	45.9	47.5	46.1	35.2	35.9	32.1	35.5	37.5	63.1	43.9	58.6	50.0	43.4

percentage weight allocated to negative options. As the complexity of the CPMP increased, the average percentage weight allocated to negative options decreased. This is evident in Table 6.3.

Table 6.3

Relationship Between Structure of CPMP and Percentage Weight Allocated to Negative Options

Complexity	CPMP	Average %
least	1	50.1
	2	48.2
	4	44.6
most	3	33.9

It seemed that as the complexity of the CPMP increased, examinees gained marks by selecting correct pathways; the largest error being that of omission rather than commission.

The descriptive statistics of the resulting examinee CPMP scores are presented in the next section.

2. Descriptive Statistics of the Examinee CPMP Scores

The mean, standard error, standard deviation, variance, kurtosis, skewness, minimum score and maximum score are presented for the proficiency, error of omission, and efficiency scores on CPMPs 1-4 in Tables 6.4 to 6.15. The distribution of these scores is presented in Appendix F. The effect that scoring procedures had upon the variance, kurtosis and skewness of proficiency scores is discussed in the following sub-sections.

A. Inference About the Variance Among Proficiency Scores Calculated Using the Twelve Scoring Procedures

A dependent sample t-test was used to test whether changes occured in the variance of scores calculated using the twelve scoring procedures. Since there were 66 paired comparisons (i.e., $(n^2 - n)/2$) of variances over all scoring procedures, the level of significance was lowered to a 0.0005 level in order to reduce the type I error which would be increased by using 66 repeated t-tests. The t-test results for CPMPs 1-4 are respectively presented in Tables 6.16 to 6.19.

With a t-critical = 3.375 (df = 109), the tables reveal that there were 48 out of 66 significant differences among the variances of CPMP 1, 41 in CPMP 2, 51 in CPMP 3, and 51 in CPMP 4. Although there was no consistent pattern over the CPMPs, there was a tendency for the Mc and CDS scoring procedures to have the lowest variance. In addition, scoring procedures with differential weighting (excluding author) tended to yield scores with larger variances. Based

Table 6.4

Descriptive Statistics of Distribution of Proficiency Scores

Calculated on CPMP 1 Using 12 Scoring Procedures

	Mean	Std. Err.[a]	Std. Dev.[b]	Variance	Kurtosis	Skewness	Min.	Max.
Author	85.357	.412	4.340	18.834	.469	- .471	70.636	84.541
GCS	81.931	.449	4.735	22.421	.295	.095	69.318	95.455
GDS	85.673	.483	5.085	25.854	.233	- .245	69.276	97.188
ICS	83.234	.465	4.899	23.997	- .119	- .104	68.831	96.104
IDS	87.488	.453	4.770	22.753	.169	- .525	73.654	98.115
ICM	87.856	.550	5.795	33.585	.502	- .644	69.568	100.000
IDM	86.423	.473	4.983	24.833	.083	- .449	72.290	98.277
CCS	84.347	.514	5.420	29.371	- .283	- .071	70.842	96.570
CDS	89.131	.334	3.519	12.382	1.740	- .995	74.928	95.533
CCM	88.167	.467	4.919	24.199	.038	- .346	74.922	98.433
CDM	89.498	.348	3.670	13.468	2.675	-1.181	74.145	95.817
Mc	86.401	.425	4.473	20.005	.352	- .254	72.292	98.448

[a] Std. Err. = Standard Error

[b] Std. Dev. = Standard Deviation

Table 6.5

Descriptive Statistics of Distribution of Error of Omission Scores

Calculated on CPMP 1 Using 12 Scoring Procedures

	Mean	Std. Err.	Std. Dev.	Variance	Kurtosis	Skewness	Min.	Max.
Author	10.317	.331	3.484	12.139	.194	.225	1.587	19.841
GCS	14.578	.392	4.134	17.090	.161	-.295	3.409	23.864
GDS	11.524	.395	4.163	17.328	-.226	.111	2.209	21.286
ICS	14.110	.423	4.452	19.825	.186	.023	2.597	25.974
IDS	10.556	.412	4.342	18.855	.161	.459	1.414	22.812
ICM	10.321	.481	5.069	25.693	.174	.471	0.000	25.062
IDM	11.218	.429	4.521	20.444	.128	.346	1.292	22.972
CCS	14.339	.488	5.145	26.466	-.019	.011	1.715	27.443
CDS	5.239	.247	2.599	6.753	.624	.623	.144	13.545
CCM	10.096	.424	4.467	19.953	.132	.148	0.000	21.943
CDM	5.046	.254	2.671	7.134	.817	.676	0.000	13.688
Mc	10.591	.377	3.968	15.747	-.110	.044	1.380	19.828

Table 6.6

Descriptive Statistics of Distribution of Efficiency Scores

Calculated on CPMP 1 Using 12 Scoring Procedures

	Mean	Std. Err.	Std. Dev.	Variance	Kurtosis	Skewness	Min.	Max.
Author	84.799	.663	6.985	48.795	-.082	-.107	65.714	100.000
GCS	87.969	.576	6.072	36.870	.265	-.367	67.500	100.000
GDS	87.969	.576	6.072	36.870	.265	-.367	67,500	100.000
ICS	87.896	.543	5.719	32.702	-.204	-.525	70.000	96.774
IDS	87.300	.498	5.243	27.489	-.701	-.409	74.286	96.774
ICM	70.783	.608	6.409	41.080	.353	-.159	51.429	86.667
IDM	86.527	.528	5.558	30.893	-.393	-.461	70.000	96.970
CCS	83.636	.539	5.678	32.238	.104	-.351	65.000	96.667
CDS	81.481	.516	5.440	29.596	.354	-.436	62.500	93.333
CCM	77.562	.540	5.692	32.400	.432	-.280	60.000	93.333
CDM	77.562	.540	5.692	32.400	.432	-.280	60.000	93.333
Mc	87.276	.583	6.138	37.675	.712	-.445	65.000	100.000

Table 6.7

Descriptive Statistics of Distribution of Proficiency Scores

Calculated on CPMP 2 Using 12 Scoring Procedures

	Mean	Std. Err.	Std. Dev.	Variance	Kurtosis	Skewness	Min.	Max.
Author	81.434	.414	4.366	19.058	- .223	- .306	67.391	90.217
GCS	80.814	.449	4.729	22.367	- .409	- .329	66.195	90.140
GDS	84.512	.461	4.855	23.567	- .453	- .328	72.856	94.286
ICS	80.568	.551	5.808	33.734	- .006	- .120	66.667	96.078
IDS	82.700	.489	5.154	26.568	.085	- .146	68.633	96.859
ICM	79.738	.541	5.704	32.540	- .036	- .037	66.037	94.339
IDM	82.057	.491	5.174	26.770	.029	- .117	67.637	96.159
CCS	86.439	.521	5.487	30.102	1.055	- .897	68.422	96.491
CDS	89.582	.393	4.144	17.170	1.814	-1.305	75.601	96.563
CCM	86.670	.488	5.143	26.455	.261	- .619	71.187	96.610
CDM	89.281	.394	4.151	17.230	1.143	-1.047	76.542	96.708
Mc	85.084	.423	4.459	19.879	- .062	- .395	70.163	93.235

Table 6.8

Descriptive Statistics of Distribution of Error of Omission Scores

Calculated on CPMP 2 Using 12 Scoring Procedures

	Mean	Std. Err.	Std. Dev.	Variance	Kurtosis	Skewness	Min.	Max.
Author	10.047	.318	3.355	11.259	.025	.320	2.174	19.565
GCS	9.479	.318	3.354	11.252	-.305	.193	2.817	16.902
GDS	10.082	.350	3.691	13.627	-.514	.294	2.857	18.572
ICS	16.058	.500	5.272	27.789	.020	.167	3.922	29.413
IDS	14.863	.461	4.853	23.554	.014	.078	2.980	27.856
ICM	16.981	.494	5.208	27.122	-.133	.099	5.661	30.189
IDM	15.393	.459	4.835	23.374	-.077	.050	3.807	28.416
CCS	12.723	.436	4.596	21.122	-.514	.241	3.509	24.561
CDS	5.108	.250	2.635	6.941	-.344	.518	.687	12.028
CCM	11.788	.422	4.446	19.764	-.638	.143	3.390	22.033
CDM	4.738	.250	2.632	6.928	-.323	.491	.412	12.346
Mc	8.945	.348	3.669	13.463	-.512	.217	1.591	17.954

Table 6.9

Descriptive Statistics of Distribution of Efficiency Scores

Calculated on CPMP 2 Using 12 Scoring Procedures

	Mean	Std. Err.	Std. Dev.	Variance	Kurtosis	Skewness	Min.	Max.
Author	78.477	.628	6.617	43.788	.031	.194	58.621	96.000
GCS	73.154	.638	6.727	45.247	-.317	.361	58.621	91.667
GDS	73.154	.638	6.727	45.247	-.317	.361	58.621	91.667
ICS	81.424	.636	6.705	44.964	-.364	.074	62.069	96.000
IDS	88.931	.467	4.922	24.225	.116	-.178	75.862	100.000
ICM	82.011	.621	6.542	42.799	-.564	.067	65.517	96.000
IDM	91.565	.380	4.005	16.042	-.217	.005	79.310	100.000
CCS	77.943	.629	6.626	43.903	.662	-.424	58.621	96.000
CDS	73.671	.626	6.596	43.510	.618	-.198	55.172	92.000
CCM	78.916	.622	6.553	42.940	.733	-.156	62.069	100.000
CDM	70.494	.634	6.681	44.634	.532	-.045	51.724	88.462
Mc	72.884	.674	7.104	50.469	-.326	.187	57.576	91.667

Table 6.10

Descriptive Statistics of Distribution of Proficiency Scores

Calculated on CPMP 3 Using 12 Scoring Procedures

	Mean	Std. Err.	Std. Dev.	Variance	Kurtosis	Skewness	Min.	Max.
Author	58.054	2.579	27.170	738.233	-1.763	.038	21.000	100.000
GCS	64.291	.866	9.129	83.341	.360	-.260	36.363	86.363
GDS	66.978	1.083	11.410	130.198	.427	-.626	30.769	89.231
ICS	53.554	1.112	11.721	137.373	.148	.335	27.778	88.889
IDS	57.487	1.012	10.661	113.661	.535	.277	29.844	89.915
ICM	55.015	1.217	12.823	164.419	.118	-.323	19.999	86.666
IDM	58.067	1.014	10.681	114.073	.481	.022	28.929	87.886
CCS	60.075	1.020	10.751	115.576	.184	.246	31.578	89.473
CDS	64.372	.908	9.563	91.444	1.213	.348	37.896	95.789
CCM	53.965	1.753	18.471	341.172	-.236	.617	17.641	100.000
CDM	53.829	1.707	17.987	323.531	-.447	.418	14.297	93.540
Mc	64.647	.883	9.304	86.562	1.309	-.055	32.810	93.301

Table 6.11

Descriptive Statistics of Distribution of Error of Commission Scores

Calculated on CPMP 3 Using 12 Scoring Procedures

	Mean	Std. Err.	Std. Dev.	Variance	Kurtosis	Skewness	Min.	Max.
Author	24.541	.309	13.258	175.775	-1.763	.038	20.000	51.278
GCS	27.478	.770	8.117	65.887	- .008	-.270	4.545	50.001
GDS	25.530	.813	8.563	73.325	- .198	.226	7.692	50.769
ICS	41.141	1.065	11.221	125.916	.249	-.537	5.536	61.111
IDS	37.979	.917	9.664	93.396	.288	-.569	8.373	57.342
ICM	42.102	1.116	11.754	138.168	- .088	-.306	6.666	66.667
IDM	37.472	.886	9.335	87.145	.289	-.543	9.772	57.827
CCS	35.610	.803	8.456	71.498	.466	-.553	10.527	52.632
CDS	24.570	.703	7.406	54.846	.356	-.816	2.105	40.000
CCM	33.249	1.723	18.152	329.513	- .851	-.368	0.000	76.476
CDM	31.903	1.630	17.175	294.981	- .756	-.106	0.000	78.561
Mc	29.413	.827	8.718	75.996	.606	-.321	2.203	55.033

Table 6.12

Descriptive Statistics of Distribution of Efficiency Scores

Calculated on CPMP 3 Using 12 Scoring Procedures

	Mean	Std. Err.	Std. Dev.	Variance	Kurtosis	Skewness	Min.	Max.
Author	67.353	1.347	14.195	201.487	- .628	.077	40.000	100.000
GCS	69.115	1.266	13.341	177.992	- .057	-.318	28.571	100.000
GDS	69.115	1.266	13.341	177.992	- .057	-.318	28.571	100.000
ICS	66.120	1.750	18.433	339.789	-1.187	-.074	30.000	100.000
IDS	82.032	.861	9.072	82.306	- .114	0.178	57.143	100.000
ICM	67.386	1.630	17.171	294.851	- .850	-.182	28.571	100.000
IDM	80.277	.887	9.349	87.398	- .354	-.107	57.143	100.000
CCS	52.211	1.307	13.770	189.625	-1.004	-.086	25.000	81.818
CDS	59.099	1.187	12.505	156.384	- .151	-.022	28.571	90.909
CCM	49.745	1.849	19.481	379.518	- .729	.053	00.000	90.909
CDM	61.893	1.656	17.443	304.256	- .313	-.584	00.000	90.909
Mc	69.115	1.266	13.341	177.992	- .057	-.318	28.571	100.000

Table 6.13

Descriptive Statistics of Distribution of Proficiency Scores

Calculated on CPMP 4 Using 12 Scoring Procedures

	Mean	Std. Err.	Std. Dev.	Variance	Kurtosis	Skewness	Min.	Max.
Author	75.400	1.013	10.672	113.885	-1.371	-.304	50.976	91.176
GCS	76.347	.099	7.364	54.223	- .140	-.447	55.085	89.792
GDS	85.248	.690	7.267	52.805	.647	-.863	60.372	98.585
ICS	79.194	.660	6.957	48.400	- .084	-.520	59.085	90.908
IDS	83.434	.657	6.921	47.896	- .560	-.186	65.832	96.915
ICM	77.379	.715	7.535	56.783	- .520	-.123	58.498	92.219
IDM	81.428	.678	7.148	51.093	- .633	-.160	63.356	94.866
CCS	79.668	.799	8.418	70.869	.044	-.532	57.149	95.239
CDS	85.647	.504	5.312	28.215	- .890	-.135	73.481	97.014
CCM	86.036	.685	7.219	52.112	- .879	-.104	69.444	100.000
CDM	84.242	.651	6.855	46.989	-1.179	-.118	71.825	96.536
Mc	82.396	.596	6.283	39.482	.223	-.470	61.562	96.686

Table 6.14

Descriptive Statistics of Distribution of Error of Commission Scores

Calculated on CPMP 4 Using 12 Scoring Procedures

	Mean	Std. Err.	Std. Dev.	Variance	Kurtosis	Skewness	Min.	Max.
Author	15.582	.661	6.969	48.562	-.823	.574	4.903	35.297
GCS	19.570	.667	7.027	48.372	.013	.698	4.083	38.790
GDS	12.504	.704	7.419	55.036	.977	1.076	0.000	38.213
ICS	18.574	.641	6.749	45.551	-.234	.574	6.819	36.369
IDS	14.197	.607	6.394	40.889	-.438	.383	3.085	30.807
ICM	20.401	.663	6.982	48.747	-.720	.144	7.781	36.315
IDM	16.090	.621	6.544	42.828	-.492	.360	5.134	32.913
CCS	19.409	.771	8.122	65.965	-.239	.444	2.380	40.470
CDS	6.793	.347	3.656	13.365	-.219	.233	.093	18.417
CCM	11.411	.563	5.926	35.121	-.523	.174	0.000	25.000
CDM	7.590	.406	4.276	18.281	-.676	.096	.231	19.630
Mc	13.949	.538	5.664	32.084	.357	.679	3.314	31.895

Table 6.15

Descriptive Statistics of Distribution of Efficiency Scores

Calculated on CPMP 4 Using 12 Scoring Procedures

	Mean	Std. Err.	Std. Dev.	Variance	Kurtosis	Skewness	Min.	Max.
Author	71.824	.555	5.849	34.214	.230	-.439	53.333	84.000
GCS	83.302	.472	4.977	24.772	-.640	-.239	72.000	93.750
GDS	83.302	.472	4.977	24.772	-.604	-.239	72.000	93.750
ICS	81.336	.617	6.497	42.217	-.374	-.188	61.538	94.737
IDS	89.411	.416	4.384	19.217	-.063	-.507	76.923	96.429
ICM	74.872	.937	9.868	97.379	-.878	-.163	55.556	94.118
IDM	90.373	.402	4.234	17.931	-.190	-.481	80.000	96.552
CCS	85.307	.636	6.701	44.897	-.404	-.472	66.667	96.296
CDS	71.123	.803	8.457	71.524	-.916	-.225	52.000	86.667
CCM	65.908	.943	9.934	98.685	-.494	.204	46.154	88.889
CDM	71.242	.835	8.793	77.321	-.672	-.195	00.000	88.889
Mc	83.036	.462	4.866	23.678	-.612	-.203	72.000	93.750

Table 6.16

T Test of Comparison of Variance Among Proficiency Scores Calculated By the Twelve Scoring Procedures on CPMP 1

AUTHOR	GCS	GDS	ICS	IDS	ICM	IDM	CCS	CDS	CCM	CDM	Mc
18.8	22.4	25.9	24.0	22.8	33.6	24.8	29.4	12.4	24.2	13.5	20.0
1.0	-3.0	-6.0*	-3.8*	-3.2	-10.9*	-5.0*	-5.9*	6.6*	-3.5*	5.5*	-1.1
	1.0	-3.8*	-1.3	-0.3	-7.8*	-2.0	-4.5*	8.3*	-1.3*	7.4*	4.3*
		1.0	1.7	3.1	-6.3*	1.1	-2.2	12.3*	1.2	11.2*	8.7*
			1.0	1.6	-8.2*	-1.1	-4.5*	10.7*	-0.2	9.3*	4.4*
				1.0	-13.4*	14.6*	-5.7*	10.2*	-1.5	8.8*	3.3
					1.0	11.4*	2.7	17.9*	7.1*	16.5*	12.4*
						1.0	-3.7*	12.3*	0.6	10.9*	5.8*
							1.0	13.1*	6.0*	11.8*	8.1*
								1.0	-10.7*	-4.4*	-8.0*
									1.0	10.0*	4.2*
										1.0	-7.1*
											1.0

* $p \leq 0.0005$

Table 6.17

T Test Comparison of
Variance Among Proficiency Scores Calculated by the Twelve
Scoring Procedures on CPMP 2

AUTHOR	GCS	GDS	ICS	IDS	ICM	IDM	CCS	CDS	CCM	CDM	Mc
19.1	22.4	23.6	33.7	26.6	32.5	26.8	30.1	17.2	26.5	17.2	19.9
1.0	-3.2	-4.0*	-8.7*	-4.6*	-7.8*	-4.6*	-5.4*	1.3	-3.9*	1.4	-0.8
	1.0	-1.3	-5.8*	-2.2	-5.2*	-2.2	-3.4*	3.4*	-1.9	3.5*	3.4*
		1.0	-5.9*	-1.8	-5.2*	-1.8	-2.8	3.9*	-1.3	3.9*	3.9*
			1.0	9.0*	2.7	7.8*	1.6	8.6*	3.3	8.7*	8.5*
				1.0	-7.6*	-0.6	-1.7	5.3*	0.1	5.3*	4.2*
					1.0	7.3*	1.0	7.9*	2.7	7.9*	7.8*
						1.0	-1.6	5.3*	0.2	5.3*	4.1*
							1.0	9.9*	4.3*	9.0*	5.6*
								1.0	-7.6*	-0.2	-2.2
									1.0	7.2*	3.8*
										1.0	-2.2
											1.0

* $p \leq 0.0005$

Table 6.18

T Test Comparison of
Variance Among Proficiency Scores Calculated by the Twelve
Scoring Procedures on CPMP 4

AUTHOR	GCS	GDS	ICS	IDS	ICM	IDM	CCS	CDS	CCM	CDM	Mc
113.9	54.2	52.8	48.4	47.9	56.8	51.1	70.9	28.2	52.1	47.0	39.5
1.0	15.6*	14.7*	14.1*	14.3*	7.9*	13.2*	7.1*	22.5*	9.6*	11.0*	20.0*
	1.0	0.7	2.3	2.3	-0.7	1.1	-4.5*	8.1*	0.5	1.6	7.7*
		1.0	1.7	1.6	1.0	0.5	-5.1*	7.5*	0.2	1.2	5.2*
			1.0	0.3	-3.4*	-1.4	-9.4*	7.1*	-1.0	0.3	4.7*
				1.0	-6.4*	-4.8*	-8.2*	8.7*	-1.6	0.3	5.5*
					1.0	4.5*	-3.7*	13.2*	1.8	3.1	8.9*
						1.0	-6.8*	-10.1*	-0.4	1.2	7.4*
							1.0	13.1*	4.4*	4.9*	12.7*
								1.0	-12.9*	-16.3*	-5.3*
									1.0	2.3	4.6*
										1.0	2.3
											1.0

* $p \leq 0.0005$

Table 6.19

T Test Comparison of
Variance Among Proficiency Scores Calculated by the Twelve
Scoring Procedures on CPMP 3

AUTHOR	GCS	GDS	ICS	IDS	ICM	IDM	CCS	CDS	CCM	CDM	Mc
738.2	83.3	130.2	137.4	113.7	164.4	114.1	115.6	91.4	341.2	323.5	86.6
1.0	48.7*	36.9*	32.4*	41.4*	30.8*	40.8*	39.3*	42.7*	9.1*	9.7*	46.7*
	1.0	-9.1	-7.7*	-5.0*	-11.0*	-5.3*	-4.3*	-1.2	-22.0*	-21.1*	-1.0
		1.0	-0.8	2.1	-4.2*	2.2	1.6	4.5*	-12.4*	-12.0*	7.0*
			1.0	6.4*	-4.6*	7.0*	2.5	5.0*	-11.0*	-10.1*	6.8*
				1.0	-9.5*	-0.3	-0.3	2.7	-14.1*	-13.1*	4.2*
					1.0	12.5*	5.7*	7.5*	-8.5*	-7.8*	10.0*
						1.0	-0.2	2.8	-14.2*	-13.2*	4.4*
							1.0	3.9*	-14.1*	-.315*	4.7*
								1.0	-20.0*	-20.9*	1.0
									1.0	1.3	21.3*
										1.0	21.1*
											1.0

* $p \leq 0.0005$

on this data, it was concluded that the method of categorization and weights assigned options could alter the variance among CPMP proficiency scores.

B. Skewness and Kurtosis

Tables 6.4 to 6.15 indicate that the skewness varied over CPMPs and scores. For example, the distribution of proficiency scores tends to be negatively skewed for CPMPs 1, 2 and 4 but positively skewed for CPMP 3. The error of omission scores tended to be positively skewed for CPMPs 1 and 4, but negatively skewed for CPMPs 2 and 3. However, the degree of skewness of scores varied over scoring procedures. For example, it was observed that the distribution of proficiency scores for CPMP 1, calculated using the scoring procedure CDM was heavily skewed to the right (_i_._e_., -1.181) but slightly skewed to the left (_i_._e_., 0.095) using the GCS scoring procedure (see Appendix F).

The kurtosis of the distribution was also altered. For example, the distribution of scores rose to a sharp point using the CDS scoring procedure (_i_._e_., kurtosis = 1.814) but was flattened using the GDS scoring procedure (_i_._e_., kurtosis = -0.409), (see Appendix F). Since no particular pattern was observed between scoring procedures and the distribution of scores, it was concluded that the distribution of scores (_i_._e_., skewness and kurtosis) could be altered by different scoring procedures.

The linear relationship among CPMP scores was determined by factor analysis and the results are presented in the next section.

3. Factor Analysis

A. Component Structure Underlying Proficiency Scores

The matrix of correlation coefficients between proficiency scores calculated using the twelve scoring procedures on CPMPs 1-4 is presented in Table 6.20. The matrix has been divided into 66 submatrices. An examination of the coefficients within the submatrices revealed the diagonal elements to be relatively large (*i.e.*, approximately 0.70) as compared to the off-diagonal elements (*i.e.*, approximately 0.10). This suggested a strong linear relationship among scores for the same CPMP regardless of scoring procedure but little relationship among scores of different CPMPs.

Table 6.21 presents the factor loading matrix from the principal component analysis. Components with eigenvalues greater than one were retained.

Factors I-IV were referred to as CPMP(1-4) test factors and had the following loadings:

CPMP 1 loaded on factor I;
CPMP 4 loaded on II;
CPMP 3 loaded on III; and
CPMP 2 loaded on IV.

Table 6.20

Correlation Coefficient* Between the Proficiency Scores of the 4 CPMPs
Calculated Using the 12 Scoring Procedures

SCORING PROCEDURE

AUTHOR				GCS				GDS				ICS				IDS				ICM				IDM				CCS				CDS				CCM				CDM				Mc			
C P M P																																															
1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
100	20	20	-09	80	15	06	-05	83	09	11	-06	74	-04	00	01	79	-07	03	01	81	-05	15	03	81	-08	06	00	59	-05	13	-04	73	10	11	03	66	-01	11	06	76	15	10	03	82	10	10	01
20	100	04	-12	31	85	03	-12	30	83	07	-11	28	69	10	-06	22	64	11	-02	21	66	10	-02	26	62	11	-03	23	42	05	-01	22	55	09	-03	25	47	01	08	20	63	02	06	32	84	04	-07
20	04	100	01	18	08	35	-02	23	-02	63	-01	19	-12	55	-02	22	-14	60	08	21	-11	69	09	20	-15	59	09	16	-01	56	-02	19	04	24	15	19	-06	-22	12	21	03	-14	19	19	01	34	04
-09	-12	01	100	-08	-10	08	84	-10	-09	07	80	-08	-07	-01	70	-06	-05	02	70	-06	-07	-02	60	-08	-05	02	60	-05	-01	00	69	-14	-08	-01	48	-08	-06	21	35	-15	-12	12	29	-09	-12	02	75
				100	30	-01	03	92	28	06	-00	85	14	03	07	83	13	04	02	82	13	12	01	85	10	06	01	77	04	15	04	61	16	05	07	78	06	00	01	66	20	-01	-09	96	26	-02	03
				30	100	01	-09	29	91	06	-08	22	65	10	-08	19	56	11	-04	16	63	08	-07	21	52	10	-06	24	36	01	-02	17	57	02	-04	23	36	-07	04	18	63	-08	03	30	93	-01	-06
				-01	01	100	06	08	04	85	13	10	-08	71	06	12	-08	71	06	12	-08	72	05	11	-08	78	06	07	-10	59	00	09	-08	60	07	05	-09	41	05	07	-08	44	06	02	-01	91	06
				03	-09	06	100	-03	-04	06	91	-01	-01	-07	85	01	01	-02	82	00	-01	-06	74	-01	02	-02	82	03	02	00	78	-11	-10	-01	43	00	-02	22	47	-10	-13	14	26	00	-10	01	90
								100	25	12	-04	89	10	08	01	90	08	10	-02	90	09	18	-03	93	05	12	03	82	02	19	-03	73	16	09	-07	82	06	01	01	76	21	00	-07	95	26	10	-01
								25	100	03	-06	23	76	05	-03	18	71	06	-01	14	75	02	-03	19	68	05	02	26	41	00	02	11	48	15	-06	23	40	-01	07	11	53	-00	01	27	89	01	-02
								12	03	100	11	13	-10	73	07	13	-10	74	08	13	-10	81	10	12	-11	78	09	08	-12	60	00	12	-07	56	11	08	-13	33	07	12	-07	42	10	08	-01	78	08
								-04	06	11	100	-02	-01	02	85	00	-02	07	76	00	-02	01	66	-01	-02	06	77	02	-01	04	79	-10	-13	04	35	01	-06	21	41	-10	-15	13	18	-01	-11	10	81
												100	11	07	04	94	08	10	-01	90	10	16	-03	95	07	11	03	88	01	21	04	72	10	13	-10	87	04	08	00	73	14	06	-12	90	21	13	00
												11	01	-03	05	04	96	-04	08	00	99	-08	06	07	95	-05	07	12	63	-03	12	-08	46	-00	01	07	64	00	14	-09	49	-02	07	10	73	-07	03
												07	-03	100	-05	07	-05	95	-01	10	-03	92	-03	07	-06	95	-01	09	-07	70	00	07	-03	48	05	09	-06	13	03	08	-03	11	06	05	04	68	-04
												04	05	-05	100	06	06	00	92	07	04	-02	87	04	07	00	92	09	07	03	90	-08	-08	02	56	07	05	22	66	-06	-09	15	43	05	-07	03	89
																100	01	10	01	95	03	17	-02	98	-01	11	-01	88	-01	19	05	75	09	10	-09	90	00	05	03	76	13	04	-10	91	17	13	02
																01	100	-05	08	-01	96	-10	06	04	99	-01	07	10	66	-03	09	-11	47	00	00	05	69	01	11	-11	48	-01	05	08	-68	-07	03
																10	-05	100	06	12	-05	91	05	10	-06	99	06	11	-02	78	03	07	02	55	11	11	-03	14	09	07	01	13	13	06	05	73	02
																01	08	06	100	00	08	02	96	-01	09	06	100	02	13	10	86	-07	-02	05	75	01	09	19	83	-07	-04	12	68	01	-04	03	93
																				100	00	18	00	96	-03	13	-01	85	-02	19	02	73	10	10	-07	87	01	04	01	75	15	02	-09	89	14	14	02
																				00	100	-08	05	06	96	-05	07	11	63	-04	10	-10	43	-03	01	06	61	-01	13	-11	45	-04	06	09	71	-09	02
																				18	-01	100	02	17	-11	95	02	16	-09	75	-02	16	03	51	08	17	-08	18	06	18	-02	22	10	14	02	70	-02
																				00	05	02	100	-03	07	05	97	00	11	09	77	-08	-02	06	80	-01	09	18	87	-06	-03	14	76	-01	-07	03	90
																								100	02	11	-02	88	00	18	02	77	10	10	-08	90	03	05	02	73	15	04	-09	92	19	13	00
																								02	100	-07	08	08	66	-03	10	-13	44	-00	01	03	68	02	12	-14	46	-00	06	06	64	-07	03
																								11	-07	100	06	11	-04	78	02	10	-01	54	11	12	-05	20	09	11	-01	20	12	08	04	75	03
																								02	08	06	100	00	13	10	86	-10	-01	05	76	-01	09	18	82	-08	-03	12	69	-01	-05	03	93
																												100	06	16	09	63	11	07	-08	94	06	02	04	65	14	-01	10	86	25	08	04
																												06	100	05	06	-08	78	07	09	02	95	-07	11	-10	73	-03	13	02	60	-06	05
																												16	05	100	03	12	05	78	14	16	06	26	14	13	04	33	17	16	02	76	06
																												09	06	03	100	-12	-11	-02	54	05	01	16	68	-11	-13	07	43	02	-03	-03	83
																																100	10	07	-11	71	-06	-12	-05	98	15	10	-07	76	12	11	-11
																																10	100	09	01	09	79	-13	-01	07	98	-08	07	15	73	-04	-08
																																01	09	100	10	04	09	43	10	04	10	58	13	07	05	84	03
																																-11	01	10	100	-11	06	11	85	-12	-01	10	94	-09	-06	06	74
																																				100	03	00	01	76	13	-02	-11	88	21	06	-01
																																				03	100	-05	08	-08	77	-01	11	05	61	-03	01
																																				00	-05	100	14	08	-12	91	06	04	-05	43	21
																																				01	08	14	100	-05	-01	10	88	01	02	04	77
																																								100	13	06	-09	80	11	09	-10
																																								13	100	-07	06	20	76	-04	-10
																																								06	-07	100	07	02	-03	51	14
																																								-09	06	07	100	-10	02	06	59
																																												100	26	05	00
																																												26	100	-03	-08
																																												05	-03	100	03
																																												00	-08	03	100

*the decimal points have been omitted

Table 6.21

Varimax Rotated Principal Component Factor Analysis of Proficiency Scores

SCORING PROCEDURE	CPMP	Factor I	II	III	IV	V	VI	VII	VIII	h_j^2
AUTHOR	1	83*	-04	02	-09	14	06	10	05	73
	2	23	-07	08	61	06	24	02	55	80
	3	18	01	63	-12	15	05	-33	03	58
	4	-09	81	01	-10	-06	00	05	03	69
GCS	1	89	03	02	13	-02	03	-04	06	81
	2	19	-06	07	55	01	21	-08	75	97
	3	03	06	80	-03	-06	-07	33	03	77
	4	-01	95	-03	-03	-11	-02	07	-01	93
GDS	1	94	-02	08	07	00	05	-02	06	91
	2	17	-02	03	72	-02	12	01	58	90
	3	06	06	83	-06	01	-07	21	09	76
	4	-02	93	05	-04	-20	-05	06	01	91
ICS	1	94	00	08	11	-05	-01	03	-03	90
	2	02	03	-05	96	03	16	00	11	97
	3	01	-03	93	02	-03	-07	-09	04	88
	4	05	94	-01	04	11	-02	07	-06	91
IDS	1	96	02	08	03	-04	00	01	-03	94
	2	00	03	-06	96	00	20	02	01	96
	3	03	03	97	02	02	00	-07	03	95
	4	00	91	03	05	38	04	02	-03	98
ICM	1	95	02	09	-01	-03	03	-01	-04	91
	2	01	02	-05	96	02	13	-02	09	96
	3	12	-03	94	-04	04	-04	-03	04	91
	4	-01	83	03	03	50	04	04	-04	94
IDM	1	98	00	07	06	-02	01	01	-02	97
	2	-02	04	-07	96	01	19	03	-02	95
	3	05	02	97	00	02	-03	-02	05	96
	4	-01	91	04	03	37	05	02	-03	98
CCS	1	87	06	07	13	-07	00	-06	-02	79
	2	-03	06	-03	55	04	75	-03	-13	88
	3	13	02	80	00	10	11	16	-13	73
	4	02	88	-01	11	13	-09	-02	-03	81
CDS	1	80	-12	05	-14	04	04	10	15	71
	2	09	-07	-01	30	02	88	-06	27	95
	3	05	-02	62	01	07	15	49	-07	66
	4	-08	54	09	-02	74	05	01	-01	86
CCM	1	91	03	07	07	-07	-01	-07	01	84
	2	-01	02	-03	56	04	77	01	-13	92
	3	04	18	17	02	-01	-10	84	00	78
	4	02	59	06	12	73	-03	02	00	90
CDM	1	83	-11	05	-15	04	02	06	15	75
	2	13	-09	-01	32	03	84	-04	33	94
	3	02	09	21	-01	03	-02	93	-01	93
	4	-09	37	11	03	90	08	-01	03	99
Mc	1	97	01	03	08	-02	03	-01	08	95
	2	16	-07	02	63	-02	42	-02	58	94
	3	05	00	81	-03	-03	01	43	-06	86
	4	00	93	00	00	-28	-02	05	-01	94
Percentage of Variance		23.10	19.40	17.20	13.70	4.60	3.80	3.10	2.50	87.40

* factor loading ≥ 45; decimal point omitted
**communality

Factors V-VII were referred to as CPMP (2-4) computer factors due to the dominant loading of the computer scoring key. The loadings under factors V-VII were as follows:

> CPMP 4, scored by the CDS, CCM and CDM procedures, loaded on V,
>
> CPMP 2, scored by the CCS, CDS, CCM and CDM procedures, loaded on VI, and
>
> CPMP 3, scored by the CCM and CDM procedures, loaded on VII.

Factor VIII was referred to as CPMP 2 group factor, with CPMP 2, scored by the GCS and Mc procedures, loading on VIII. Factor VIII was referred to as CPMP 2 group factor since the group method of categorization is used in both the GCS and Mc scoring procedures. Because this method of categorization was common for the Mc, GCS and GDS scoring procedures, factors on which Mc loaded with the GCS and GDS methods, will be referred to as 'group'.

The percentage of variance accounted for by each factor is presented at the bottom of Table 6.21. In total, the eight factors accounted for 87.4% of the observed score variance. Of this, CPMP (1-4) test factors accounted for 73.4%; CPMP (2-4) computer factors, 11.5%; and CPMP 2 group factor, 2.5%.

The above analytical results suggested a predominant clinical problem factor (factors I-IV), and a minor scoring factor (factors V-VIII). Due to the scoring factors, each clinical problem was further factor analyzed to determine the scoring structure within each CPMP.

B. Component Analysis of Proficiency Scores on Each CPMP

Component analysis of scores for CPMPs 1 to 4 was undertaken as follows:

1) correlation coefficients between proficiency scores on each CPMP were calculated using the twelve scoring procedures, and

2) the correlation coefficient matrix was subjected to the same principal component analysis previously applied to the 48 x 48 correlation matrix.

a) CPMP 1

Table 6.22 presents the correlation coefficients between proficiency scores on CPMP 1.

Table 6.23 presents the resulting matrix from the principal component analysis.

In the principal components factor analysis, one component was found to underlie the correlation matrix accounting for 84.1% of the observed variance. This supports results illustrated in Table 6.20 where all scoring methods loaded highly on factor 1 for CPMP 1.

Table 6.22

Correlation Coefficient* Between Proficiency Scores on CPMP 1
Calculated Using the 12 Scoring Procedures (N = 111)

	AUTHOR	GCS	GDS	ICS	IDS	ICM	IDM	CCS	CDS	CCM	CDM	Mc
AUTHOR	100	80	83	74	79	81	81	59	73	66	76	82
GCS		100	92	85	83	83	85	77	61	78	66	96
GDS			100	89	90	90	93	80	73	82	76	95
ICS				100	94	90	95	88	72	87	73	90
IDS					100	95	98	88	75	90	76	91
ICM						100	96	85	73	87	75	89
IDM							100	88	77	90	78	92
CCS								100	63	94	65	86
CDS									100	71	98	76
CCM										100	76	88
CDM											100	80
Mc												100

*decimal point omitted

Table 6.23

Principal Component Factor Analysis
of Proficiency Scores on CPMP 1

Scoring Procedure	Factor I	h_j^2**
Author	83*	68
GCS	89	79
GDS	95	90
ICS	94	89
IDS	97	94
ICM	95	90
IDM	98	97
CCS	88	77
CDS	80	64
CCM	91	83
CDM	83	69
Mc	97	94
Percentage of Variance	84.10	84.10

*decimal point omitted

**communality

b) CPMP 2

Table 6.24 presents the correlation coefficients between proficiency scores on CPMP 2.

Table 6.24

Correlation Coefficient* Between Proficiency Scores on CPMP 2
Calculated Using the 12 Scoring Procedures (N = 111)

	AUTHOR	GCS	GDS	ICS	IDS	ICM	IDM	CCS	CDS	CCM	CDM	Mc
AUTHOR	100	85	83	69	64	66	62	40	55	47	63	84
GCS		100	91	65	56	63	52	36	57	36	63	93
GDS			100	76	71	75	68	41	49	40	53	89
ICS				100	96	99	95	64	46	64	49	73
IDS					100	96	99	66	47	69	48	68
ICM						100	96	63	43	61	45	71
IDM							100	66	44	68	45	64
CCS								100	78	95	73	60
CDS									100	79	98	73
CCM										100	77	61
CDM											100	73
Mc												100

* decimal point omitted

Table 6.25 presents the resulting matrix from the principal component analysis.

Table 6.25

Varimax Rotated Principal Component Factor Analysis
of Proficiency Scores on CPMP 2

Scoring Procedure	Factor I	Factor II	Factor III	h_j^2**
Author	36	76*	26	78
GCS	24	94	20	97
GDS	46	82	14	90
ICS	86	42	24	97
IDS	89	32	28	98
ICM	87	39	21	97
IDM	91	28	27	97
CCS	49	07	80	89
CDS	10	39	89	95
CCM	49	08	83	93
CDM	10	47	84	94
Mc	34	80	43	95
Percentage of Variance	34.42	30.58	28.25	93.25

*factor loading ≥ 45; decimal point omitted
**communality

In the above analysis, three components were found to underlie the correlation matrix. These rotated components (*i.e.*, factors) were related to the methods of categorizing options. Categorization of options by individual judgement loaded on factor I; by author and group on II; and by computer on III. Factors I, II and III accounted respectively for 34.2%, 30.58%, and 28.25% of the total observed variance of 93.25%. It would appear that three separate components are produced by group, individual and computer scoring procedures.

c) CPMP 3

Table 6.26 presents the correlation coefficients between proficiency scores on CPMP 3.

Table 6.26

Correlation Coefficient* Between Proficiency Scores on CPMP 3 Calculated Using the 12 Scoring Procedures (N = 111)

	AUTHOR	GCS	GDS	ICS	IDS	ICM	IDM	CCS	CDS	CCM	CDM	Mc
AUTHOR	100	35	63	55	60	69	59	56	24	-22	-14	34
GCS		100	85	71	75	72	78	59	60	41	44	91
GDS			100	73	74	81	78	60	56	33	42	78
ICS				100	95	92	96	70	48	13	11	68
IDS					100	91	99	79	55	14	13	73
ICM						100	95	75	51	18	22	70
IDM							100	78	55	20	20	75
CCS								100	78	26	33	76
CDS									100	43	58	84
CCM										100	91	43
CDM											100	51
Mc												100

*decimal point omitted

Table 6.27 presents the resulting matrix from principal component analysis.

Table 6.27

Varimax Rotated Principal Component Factor Analysis of Proficiency Scores on CPMP 3

Scoring Procedure	Factor		h_j^{2**}
	I	II	
Author	70*	-21	53
GCS	73	47	75
GDS	79	35	74
ICS	92	08	86
IDS	96	11	94
ICM	94	14	91
IDM	96	17	95
CCS	76	32	68
CDS	52	58	61
CCM	02	87	75
CDM	04	97	94
Mc	72	57	85
Percentage of Variance	54.80	24.40	79.30

*factor loading ≥ 45; decimal point omitted
**communality

In the above analysis, two components were found to underlie the scoring procedures on CPMP 3. The rotated components were again found to be related to methods of categorizing options: author, group and individual loaded on factor I and computer on II. Factors I and II respectively accounted for 54.8% and 24.4% of the total observed variance of 79.3%.

d) CPMP 4

Table 6.28 presents the correlation coefficients between proficiency scores on CPMP 4.

Table 6.28

Correlation Coefficient* Between Proficiency Scores on CPMP 4
Calculated Using the 12 Scoring Procedures (N = 111)

	AUTHOR	GCS	GDS	ICS	IDS	ICM	IDM	CCS	CDS	CCM	CDM	Mc
AUTHOR	100	84	80	70	70	60	71	69	48	35	29	75
GCS		100	91	85	82	74	82	78	43	47	26	90
GDS			100	85	76	66	77	79	35	41	18	81
ICS				100	92	87	92	90	56	66	43	89
IDS					100	96	100	86	75	83	68	93
ICM						100	97	77	80	86	76	90
IDM							100	86	76	82	69	93
CCS								100	54	68	43	83
CDS									100	85	94	74
CCM										100	88	77
CDM											100	59
Mc												100

*decimal point omitted

Table 6.29 presents the resulting matrix from the principal component analysis.

Table 6.29

Varimax Rotated Principal Component Factor Analysis of Proficiency Scores on CPMP 4

Scoring Procedure	Factor I	Factor II	h_j^2**
Author	79*	19	66
GCS	95	18	93
GDS	95	10	91
ICS	86	40	90
IDS	75	64	98
ICM	63	73	94
IDM	75	64	98
CCS	79	40	79
CDS	28	88	86
CCM	34	88	89
CDM	07	99	98
Mc	80	54	94
Percentage of Variance	51.80	38.10	89.50

*factor loading ≥ 45; decimal point omitted
**communality

Two components were found to underlie the scoring procedures on CPMP 4. Factor loadings were again related to the method of categorizing options: author and group loaded on factor I, and computer on II. The individual method loaded on both I and II. Factors I and II respectively accounted for 51.8% and 38.1% of the total observed variance of 89.5%.

C. Discussion of the Component Analytic Investigation of Proficiency Scores

A component analysis of the 48 x 48 correlation matrix resulted in eight factors which when interpreted were given the following names:

Factor I:	CPMP 1 factor
Factor II:	CPMP 4 factor
Factor III:	CPMP 3 factor
Factor IV:	CPMP 2 factor
Factor V:	CPMP 4, computer factor
Factor VI:	CPMP 2, computer factor
Factor VII:	CPMP 3, computer (multiple key) factor
Factor VIII:	CPMP 2, group factor

It was observed that more than one scoring procedure, but only one CPMP loaded on each factor. This observation suggested that performance on different CPMPs was linearly unrelated ($r \doteq 0$). Thus, irrespective of the scoring procedure, there was little linear relationship between student performance on different simulated problems. Proficiency, as measured by the computer simulations, was case specific.

Varying scoring procedures did alter the linear relationship of proficiency scores within, but not across, cases. This alteration was observed in the loadings of the last four factors in Table 6.21 which accounted for a small, but significant, amount of the observed score variance (14.0%). In order to further understand this observed alteration, the scores of each CPMP were factor analyzed. In this analysis, no consistent relationship was observed between the CPMPs and the number of factors: one factor was observed in CPMP 1, three in CPMP 2, two in CPMP 3, and, two in CPMP 4. A relationship did exist, however, between the loadings on each component and the methods used to categorize options within scoring procedures, but this relationship was not consistent over CPMPs as illustrated below:

CPMP 1:

Factor I: author, group, individual and computer

CPMP 2:

Factor I: author and group
Factor II: individual
Factor III: computer

CPMP 3:

Factor I: author, group and individual
Factor II: computer

CPMP 4:

Factor I: author,group and individual
Factor II: computer, and individual

It therefore appeared that both simulated clinical problems and categorization methods determined the linear relationship among proficiency scores.

D. Component Structure Underlying Error of Commission Scores

The correlation coefficients between error of commission scores for the four CPMPs are presented in Table 6.30. A pattern of coefficients within the 66 submatricies was observed which was similar to that of the proficiency score coefficients. The relatively large diagonal coefficients and smaller off-diagonal coefficients suggested little relationship between scores of different CPMPs.

Table 6.31 presents the factor loading matrix from the principal component analysis. Factors I - IV in Table 6.31 had the following loadings:

CPMP 1, loaded on factor I,

CPMP 2, loaded on II,

CPMP 4, loaded on III, and

CPMP 3, loaded on IV.

The remaining three factors had the following loadings:

CPMP 3, scored by the CCM and CDM methods, loaded on factor V,

CPMP 4, scored by the author, GCS and GDS methods, loaded on VI, and

CPMP 2, scored by the CCS, CDS, CCM and CDM methods, loaded on VII.

Table 6.30

Correlation Coefficient* Between Errors of Commission Calculated Using
12 Scoring Procedures and 4 CPMPs

SCORING PROCEDURE

C P M P

AUTHOR				GCS				GDS				ICS				IDS				ICM				IDM				CCS				CDS				CCM				CDM				Mc			
1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
100	08	10	-03	88	07	00	05	94	09	05	01	84	03	08	11	91	01	06	06	89	02	09	07	92	00	05	05	85	02	10	10	82	00	02	07	85	01	-03	16	83	04	-05	05	88	11	04	11
08	100	-05	-03	20	91	-03	02	12	87	-10	00	15	92	-05	10	06	93	-02	12	02	92	-06	10	11	93	-03	11	16	76	-02	13	07	70	07	06	15	81	04	17	06	75	07	11	16	89	-01	07
10	-05	100	-09	09	-07	-29	-04	04	-06	-21	-08	06	-03	-25	-01	08	-03	-32	-03	03	-02	-22	-01	06	-03	-31	-03	05	-06	-42	-02	07	-08	-40	00	11	-04	-18	-01	08	-07	-21	00	09	-09	-39	-01
-03	-03	-09	100	01	02	-04	91	-06	02	-03	88	00	00	-06	79	-02	-01	-05	75	-02	-01	-08	66	-03	00	-06	76	01	01	-11	81	-05	-05	-05	70	00	-02	11	52	-06	-07	05	57	-02	-02	-09	79
				100	20	-02	11	92	21	01	05	92	15	10	15	89	14	08	07	88	14	09	06	91	12	06	05	92	10	10	14	82	08	03	05	90	10	-08	12	82	13	-08	00	96	24	00	13
				20	100	-06	06	11	96	-09	06	15	91	-04	12	06	89	-02	13	01	91	-07	09	09	88	-04	11	15	77	-04	13	08	72	-01	08	14	76	02	17	08	75	06	12	16	95	-07	10
				-02	-06	100	-04	00	-07	66	06	03	-07	77	-04	02	-08	81	-02	04	-08	81	-04	02	-08	87	-02	02	-06	73	00	02	-09	74	-05	02	-04	21	-03	03	-09	22	-03	00	-05	93	-04
				11	06	-04	100	03	07	-05	93	08	06	-08	89	07	06	-08	02	07	05	-09	73	06	06	-08	82	10	04	-11	86	04	-04	-04	68	08	00	16	62	04	-04	10	57	08	02	-08	89
								100	11	06	-02	89	06	10	07	94	06	09	00	94	06	10	00	97	04	07	-01	88	03	10	05	86	03	02	-02	89	03	-06	07	86	07	-06	-05	93	14	03	05
								11	100	-11	06	14	87	-07	11	08	87	-05	12	02	87	-10	08	10	86	-07	11	15	72	-06	13	09	66	00	06	15	70	00	16	09	69	05	10	17	91	-08	10
								06	-11	100	00	03	-15	77	-07	04	-17	77	-03	05	-15	80	-01	04	-18	80	-01	02	-12	63	-05	03	-14	27	-04	02	-14	04	-03	03	-16	11	-01	03	-12	62	-05
								-02	06	00	100	04	05	-01	86	03	03	00	76	02	04	-02	64	02	03	00	76	05	01	-05	80	01	-06	05	61	05	-03	16	53	01	-07	12	51	04	00	01	80
												100	11	12	14	93	09	10	07	90	10	11	06	94	08	09	05	96	06	13	14	90	04	08	05	94	06	-01	11	90	08	-03	01	95	20	04	11
												11	100	-09	13	02	97	-08	15	-03	99	-12	12	05	96	-09	14	11	76	-06	15	04	70	02	09	10	77	03	20	03	74	05	14	10	86	-07	11
												12	-09	100	-07	12	-11	98	00	14	-09	96	-02	11	-12	97	00	12	-05	88	00	10	-08	57	01	12	-06	-08	-01	11	-09	-08	03	12	-04	71	-05
												14	13	-07	100	13	02	-07	91	13	12	-06	85	12	12	-07	90	15	-11	-11	92	10	05	-05	77	14	10	11	76	09	05	08	70	14	10	-08	92
																100	00	10	05	96	01	12	04	99	-01	09	04	94	-01	12	13	93	-03	06	02	94	-01	-03	10	93	01	-06	-01	95	11	04	09
																00	100	-10	13	-04	97	-14	10	04	99	-11	12	09	79	-08	13	02	74	02	06	08	82	02	17	01	77	05	11	09	84	-07	10
																10	-10	100	-01	12	-08	96	-02	10	-11	99	-01	10	-04	86	-01	08	-06	62	-03	09	-05	-07	-02	09	-08	-07	02	10	-02	75	-06
																05	13	-01	100	04	14	-01	97	04	14	-02	100	08	13	-04	91	02	07	-05	89	07	10	04	91	03	06	00	89	05	09	-05	94
																				100	-04	14	05	97	-05	11	03	89	-04	15	11	87	-04	07	04	89	-03	-02	11	87	-01	-04	00	91	07	06	10
																				-04	100	-13	11	04	96	-10	13	10	76	-07	14	02	07	00	08	08	76	01	18	02	74	02	13	09	86	-08	10
																				14	-13	100	-01	11	-15	97	-01	11	-10	82	-02	11	-13	52	-02	11	-10	01	00	12	-14	02	02	11	-08	72	-06
																				05	11	-01	100	04	11	-03	97	07	11	-04	83	01	06	-08	91	06	09	03	93	02	05	01	93	05	06	-06	91
																								100	02	08	03	94	02	12	11	92	00	05	02	95	02	-02	11	92	05	-04	-01	96	14	04	08
																								02	100	-12	12	08	79	-08	14	00	73	03	06	07	81	03	17	-01	77	06	11	07	-83	-07	10
																								08	-12	100	-02	09	-06	86	-02	07	-09	63	-03	08	-07	04	-02	08	-10	05	01	08	-04	79	-06
																								03	12	-02	100	06	12	-03	91	01	06	-05	90	05	10	04	90	01	05	00	90	04	08	-05	94
																												100	09	13	16	91	06	07	06	97	09	01	12	91	11	-03	02	97	21	03	12
																												09	100	-01	11	03	93	06	10	06	97	-06	15	01	92	-02	13	06	82	-03	08
																												13	-01	100	-04	13	-02	74	-05	12	-01	-01	-01	14	-04	01	00	12	-01	81	-08
																												16	11	-04	100	10	02	[illegible]	81	15	09	08	78	10	01	02	74	12	10	-05	92
																																100	00	08	-02	94	01	06	07	99	04	03	-04	92	14	04	05
																																00	100	02	05	02	92	-07	11	-02	99	-04	10	04	80	-05	00
																																08	02	100	-09	07	08	14	-07	07	03	16	-07	05	05	84	-07
																																-02	05	-09	100	03	08	-01	90	-02	04	-05	96	02	07	-10	90
																																				100	06	02	10	95	07	-01	00	97	19	03	09
																																				06	100	-03	14	-01	92	00	11	06	82	-01	05
																																				02	-03	100	03	04	-04	93	-04	-02	04	14	08
																																				10	14	03	100	07	11	02	94	10	14	-05	88
																																								100	02	00	-04	93	12	05	04
																																								02	100	-01	08	09	84	-06	00
																																								00	-01	100	-07	-04	07	17	04
																																								-04	08	-07	100	-02	10	-06	84
																																												100	21	02	10
																																												21	100	-04	06
																																												02	-04	100	-08
																																												10	06	-08	100

*decimal points have been omitted

Table 6.31

Varimax Rotated Principal Component Factor Analysis of Error of Commission Scores

SCORING PROCEDURE	CPMP	FACTOR I	II	III	IV	V	VI	VII	h_j^2**
AUTHOR	1	91*	00	05	01	-03	-08	01	84
	2	08	95	04	01	04	-03	-08	91
	3	-09	-06	-03	-36	-18	-09	-12	20
	4	-04	-05	80	-04	04	43	01	84
GCS	1	93	13	05	01	-08	04	-04	89
	2	07	95	06	-01	02	02	-11	92
	3	-01	-04	-01	89	19	04	00	84
	4	05	-01	86	-06	09	46	-03	96
GDS	1	95	04	-02	02	-05	-01	-01	91
	2	08	91	05	-04	01	04	-15	86
	3	02	-11	-01	75	-01	-06	-13	59
	4	00	-02	79	03	10	52	-05	91
ICS	1	95	08	05	04	-01	01	-01	92
	2	03	96	07	-05	03	-01	-14	95
	3	09	-05	00	94	-17	-05	-08	94
	4	10	06	91	-05	07	22	01	90
IDS	1	98	-01	03	03	-03	01	-01	97
	2	01	96	06	-06	03	01	-07	94
	3	06	-03	-01	97	-15	-02	-07	97
	4	02	08	98	00	-01	-04	00	97
ICM	1	95	-06	04	06	-01	-02	04	91
	2	02	96	06	-06	01	00	-14	95
	3	09	-08	-01	93	-07	-10	-13	91
	4	02	05	95	-01	-01	-20	03	95
IDM	1	99	02	02	03	-01	-01	00	98
	2	00	96	06	-07	04	01	-07	93
	3	05	-05	-01	98	-03	-04	-08	98
	4	01	07	99	00	-01	-04	01	98
CCS	1	96	09	06	04	00	02	02	93
	2	00	86	06	00	06	00	42	92
	3	10	-02	-04	90	-04	-05	09	83
	4	09	07	92	00	02	17	-04	89
CDS	1	94	01	00	03	07	00	03	89
	2	-01	82	00	-03	-08	-05	52	94
	3	04	05	-05	70	16	12	17	57
	4	00	03	92	-03	-06	-20	04	90
CCM	1	96	07	04	03	02	02	-02	95
	2	00	87	04	00	03	-03	41	92
	3	-01	00	05	05	92	04	-02	86
	4	08	12	89	01	01	-35	00	95
CDM	1	94	00	00	04	05	00	00	89
	2	03	85	-01	-04	-05	-05	47	94
	3	-03	04	01	08	97	-01	00	95
	4	-04	08	89	01	-08	-37	02	95
Mc	1	98	08	03	03	-02	03	-03	97
	2	13	92	03	00	04	00	07	88
	3	01	03	-05	87	15	05	11	80
	4	07	03	97	-05	03	06	-03	96
Percentage of Variance		22.98	20.96	20.73	15.09	4.25	2.41	2.21	88.63

*factor loading ≥ 45; decimal point omitted

**commonality

Based upon the above factor loadings, the following names were attached to factors I-VII: CPMP 1 factor, CPMP 2 factor, CPMP 4 factor, CPMP 3 factor, CPMP 3 computer (multiple key) factor, CPMP 4 author and group factor, and CPMP 2 computer factor.

The percentage of variance accounted for by each factor is presented at the bottom of Table 6.31. In total, the seven factors accounted for 88.6% of the observed score variance. Of this, the CPMP 1-4 factors accounted for 79.8% and the last three factors accounted for 8.8%.

The above analytical results suggested a dominant clinical problem factor (factors I-IV) and a minor scoring factor (factors V-VII). To examine the affect that scoring procedures could have upon the relationship of commission scores, each CPMP was further factor analyzed.

E. Component Analysis of Commission Scores on Each CPMP

The procedure for component analysis of commission scores on CPMPs 1-4 was the same as that undertaken for proficiency scores (see page 129).

a) CPMP 1

Table 6.32 presents the correlation coefficients between error of commission scores on CPMP 1.

Table 6.33 presents the resulting matrix from the principal component analysis.

Table 6.32

Correlation Coefficient* Between Errors of Commission on CPMP 1 Calculated Using the 12 Scoring Procedures (N = 111)

	AUTHOR	GCS	GDS	ICS	IDS	ICM	IDM	CCS	CDS	CCM	CDM	Mc
AUTHOR	100	88	94	84	91	89	92	85	82	85	83	88
GCS		100	92	92	89	88	91	92	82	90	82	96
GDS			100	89	94	94	97	88	86	89	86	93
ICS				100	93	90	94	96	90	94	90	95
IDS					100	96	99	94	93	94	93	95
ICM						100	97	89	87	89	87	91
IDM							100	94	92	95	92	96
CCS								100	91	97	91	97
CDS									100	94	99	92
CCM										100	95	97
CDM											100	93
Mc												100

*decimal point omitted

Table 6.33

Principal Components Factor Analysis of Errors of Commission Scores on CPMP 1

Scoring Procedure	Factor I	h_j^2**
Author	98	97
GCS	93	87
GDS	95	90
ICS	96	92
IDS	98	97
ICM	95	90
IDM	99	98
CCS	96	93
CDS	94	88
CCM	97	94
CDM	94	89
Mc	98	97
Percentage of Variance	92.33	

*decimal point omitted
**communality

In the above analysis, one component was found to underlie the correlation matrix, accounting for 92.33% of the observed variance. All scoring methods load highly on the single factor.

b) CPMP 2

Table 6.34 presents the correlation coefficients between error of commission scores on CPMP 2.

Table 6.34

Correlation Coefficient* Between Errors of Commission on CPMP 2
Calculated Using the 12 Scoring Procedures (N = 111)

	AUTHOR	GCS	GDS	ICS	IDS	ICM	IDM	CCS	CDS	CCM	CDM	Mc
AUTHOR	100	91	87	92	93	92	93	76	70	81	75	89
GCS		100	96	91	89	91	88	77	72	76	75	95
GDS			100	87	87	87	86	72	66	70	69	91
ICS				100	97	99	96	76	70	77	74	86
IDS					100	97	99	79	74	82	77	84
ICM						100	96	76	70	76	74	86
IDM							100	79	73	81	77	83
CCS								100	93	97	92	82
CDS									100	92	99	80
CCM										100	92	82
CDM											100	84
Mc												100

*decimal point omitted

Table 6.35 presents the resulting matrix from the principal component analysis.

Table 6.35

Varimax Rotated Principal Component Factor Analysis of Errors of Commission Scores on CPMP 2

Scoring Procedure	Factor I	Factor II	h_j^2**
Author	84*	45	91
GCS	85	44	91
GDS	84	38	85
ICS	89	40	96
IDS	86	46	94
ICM	89	40	95
IDM	85	46	93
CCS	47	84	93
CDS	37	91	97
CCM	49	83	92
CDM	43	88	95
Mc	73	57	87
Percentage of Variance	54.13	38.38	92.51

*factor loading ≥ 45; decimal point omitted
**communality

Two components were found to underlie the correlation matrix. These rotated components were related to the methods of categorizing options. Categorization by author, group and individual methods loaded on factor I, and by the computer method on II. Factors I and II accounted respectively for 54.13% and 38.38% of the total observed variance of 92.15%.

c) CPMP 3

Table 6.36 presents the correlation coefficients between error of commission scores on CPMP 3.

Table 6.36

Correlation Coefficient* Between Errors of Commission on CPMP 3
Calculated Using the 12 Scoring Procedures (N = 111)

	AUTHOR	GCS	GDS	ICS	IDS	ICM	IDM	CCS	CDS	CCM	CDM	Mc
AUTHOR	100	-29	-21	-25	-32	-22	-31	-42	-40	-18	-21	-39
GCS		100	66	77	81	81	87	73	74	21	22	93
GDS			100	77	77	80	80	63	27	04	11	62
ICS				100	98	96	97	88	57	-08	-08	71
IDS					100	96	99	86	62	-07	-07	75
ICM						100	97	82	52	01	02	72
IDM							100	86	63	04	05	79
CCS								100	74	-01	01	81
CDS									100	14	16	84
CCM										100	93	14
CDM											100	17
Mc												100

*decimal point omitted

Table 6.37 presents the resulting matrix from the principal component analysis.

Table 6.37

Varimax Rotated Principal Component Factor Analysis of Errors of Commission Scores on CPMP 3

Scoring Procedure	Factor I	II	III	h_j^2**
Author	-17	-37	-15	19
GCS	67*	59	18	83
GDS	83	12	09	71
ICS	91	34	-11	96
IDS	89	40	-10	97
ICM	94	28	00	96
IDM	91	41	02	100
CCS	69	59	-06	83
CDS	25	96	05	100
CCM	-03	10	93	87
CDM	-01	11	99	100
Mc	56	74	11	87
Percentage of Variance	44.60	23.84	16.24	84.68

*factor loading ≥ 45; decimal point omitted
**communality

Three components were found to underlie the scoring procedures of CPMP 3. Factor loadings were again related to the method of categorizing options: group and individual methods loaded on factor I, computer (single key) on II, computer (multiple key) on III, and McLaughlin and CCS on I and II. It is interesting to note that only 19% of the variance was accounted for in error of commission scores generated using the author's key. This is not surprising since this key had very few negative options.

Factors I, II and III respectively accounted for 44.60%, 23.84% and 16.24% of the total observed variance of 84.68%.

d) CPMP 4

Table 6.38 presents the correlation coefficients between error of commission scores on CPMP 4.

Table 6.39 presents the resulting matrix from the principal component analysis.

Two components were found to underlie the scoring procedures on CPMP 4. The factors appeared to be defined by the method of categorizing options, but the pattern was not obvious. All scoring procedures except the author and group methods loaded on factor I, and all procedures except CDS, CCM and CDM loaded on II. Factors I and II respectively accounted for 49.23% and 42.86% of the total observed score variance of 92.96%.

Table 6.38

Correlation Coefficient* Between Errors of Commission on CPMP 4 Calculated Using the 12 Scoring Procedures (N = 111)

	AUTHOR	GCS	GDS	ICS	IDS	ICM	IDM	CCS	CDS	CCM	CDM	Mc
AUTHOR	100	91	88	79	75	66	76	81	70	52	57	79
GCS		100	93	89	82	73	82	86	68	62	57	89
GDS			100	86	76	64	76	80	61	53	51	80
ICS				100	91	85	90	92	77	76	70	92
IDS					100	97	100	91	89	91	89	94
ICM						100	97	83	91	93	93	91
IDM							100	91	90	90	90	94
CCS								100	81	78	74	92
CDS									100	90	96	90
CCM										100	94	88
CDM											100	84
Mc												100

*decimal point omitted

Table 6.39

Varimax Rotated Principal Component Factor Analysis of Error of Commission Scores on CPMP 4

Scoring Procedure	Factor I	Factor II	h_j^2**
Author	34	84*	84
GCS	36	92	97
GDS	28	91	91
ICS	57	76	90
IDS	79	59	97
ICM	87	45	96
IDS	79	59	98
CCS	61	71	88
CDS	84	42	89
CCM	92	30	94
CDM	94	26	95
Mc	72	67	96
Percentage of Variance	49.83	42.86	92.69

*factor loading ≥ 45; decimal point omitted
**communality

F. Discussion of the Factor Analytic Investigation of Error of Commission Scores

A component analysis of the 48 x 48 correlation matrix resulted in seven factors which when interpreted were given the following names:

Factor I: CPMP 1 factor
Factor II: CPMP 2 factor
Factor III: CPMP 4 factor
Factor IV: CPMP 3 factor
Factor V: CPMP 3 computer (multiple key) factor
Factor VI: CPMP 4, author and group factor
Factor VII: CPMP 2, computer factor

It was observed that more than one scoring procedure, but only one CPMP, loaded on each factor. This observation suggested that performance on different CPMPs was not linearly related ($r \doteq 0$). Thus, irrespective of the scoring procedure, there was little linear relationship among scores of different simulated problems. Error of commission, as measured by the computer simulations, was case specific.

Varying scoring procedures did alter the linear relationship of error of commission scores within, but not across, cases. This alteration was observed in the last three factors in Table 6.31 which accounted for 8.8% of the observed score variance.

Each CPMP was factor analyzed to determine the linear relationship of error of commission scores within each problem. No consistent relationship was observed between the CPMPs and the number of factors: one factor was observed in CPMP 1, two in CPMP 2, three in CPMP 3, and two in CPMP 4. A relationship did exist, however, between methods used to categorize options within scoring procedures and loadings on each factor, but this relationship was not consistent over CPMPs as illustrated below:

CPMP 1:

Factor I: author, group individual and computer

CPMP 2:

Factor I: author, group and individual
Factor II: computer

CPMP 3:

Factor I: group and individual
Factor II: computer

CPMP 4:

Factor I: computer and individual
Factor II: author, group and individual

It therefore appeared that both simulated clinical problems and categorization methods determined the linear relationship among error of commission scores.

G. Component Structure Underlying Error of Omission Scores

The correlation coefficients between error of omission scores for the four CPMPs are presented in Table 6.40. The pattern of coefficients within the 66 submatrices was similar to that found in the proficiency and error of commission scores.

Table 6.41 presents the factor loading matrix from the principal component analysis. Nine factors were found to underlie the correlation matrix. The following factor loadings were dependent upon both the CPMPs and the scoring procedures:

CPMP 1 and all scoring procedures loaded on factor I,

CPMP 4 and the author, GCS, GDS, ICS, IDS, ICM, IDM, CDS and Mc methods loaded on factor II,

CPMP 3 and the GCS, GDS, ICS, IDS, ICM, IDM, CDS and Mc methods loaded on factor III,

CPMP 2 and the ICS, IDS, ICM, and IDM methods loaded on factor IV,

CPMP 2 and the author, GCS, GDS, CCS, CDM and Mc methods loaded on V,

CPMP 2 and the IDM, CCS and CDS methods loaded on factor VI,

CPMP 4 and the CDS, CCM and CDM methods loaded on factor VII,

CPMP 3 and the CCM and CDM methods loaded on factor VIII, and

CPMP 3 and the author, ICS and CCS methods loaded on factor IX.

Table 6.40

Correlation Coefficient* Between Errors of Omission Calculated Using
12 Scoring Procedures and 4 CPMPs

SCORING PROCEDURE

AUTHOR				GCS				GDS				ICS				IDS				ICM				IDM				CCS				CDS				CCM				CDM				Mc			
																								C P M P																							
1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
100	35	08	00	87	26	05	14	72	12	01	13	75	16	07	14	79	17	03	15	73	20	16	17	82	16	10	15	45	10	-08	-11	85	26	14	19	55	22	05	-01	86	30	-04	19	87	30	04	13
35	100	21	-04	40	90	09	-02	43	78	06	-02	37	68	12	-09	40	66	13	-05	41	68	20	-05	39	63	19	-06	32	37	-09	00	40	74	10	02	34	51	-05	01	40	80	-09	07	41	88	09	-02
08	21	100	07	14	19	28	13	20	09	24	15	19	13	-12	11	19	17	22	16	20	15	37	09	20	17	40	16	16	14	42	11	12	15	42	19	19	14	-05	20	14	15	-05	23	13	19	31	16
00	-04	07	100	-04	-07	13	57	-04	-07	14	49	-04	-06	04	56	-03	03	10	60	-02	-05	06	40	-04	03	09	59	-10	07	09	20	-03	02	09	32	-11	03	12	12	-06	-02	20	19	-03	-07	13	66
				100	34	07	11	84	21	01	06	81	23	03	10	79	23	03	12	70	27	16	15	81	22	11	11	51	10	-03	-16	80	29	14	11	55	22	04	70	83	33	-03	10	99	37	06	08
				34	100	10	-03	38	84	04	-08	31	58	12	-15	36	53	14	-09	36	58	17	-09	34	49	19	-09	37	29	-08	06	34	74	10	00	34	40	-05	01	34	79	-04	05	37	97	09	-04
				07	10	100	19	09	12	85	16	13	-01	43	13	18	03	63	17	16	00	72	11	15	01	62	17	10	-06	41	-06	06	02	62	03	01	-09	32	-02	04	01	31	03	01	09	97	18
				11	-03	19	100	09	-05	21	78	08	-03	03	71	10	06	13	75	12	-01	15	67	12	06	16	74	01	11	17	18	15	18	25	55	04	13	18	31	13	08	23	43	12	-01	20	97
								100	29	06	04	77	26	03	03	81	27	08	03	75	30	17	05	82	26	16	03	63	09	08	-18	74	30	15	08	63	22	03	-06	76	34	-03	10	88	41	09	05
								29	100	09	-08	23	61	16	-23	25	62	18	-16	28	61	18	-19	24	57	19	-17	32	24	-04	-02	22	61	11	-10	26	29	-03	-03	21	64	-02	-04	25	82	13	-07
								06	09	100	22	11	-07	44	14	12	-03	68	19	12	-06	66	12	11	-05	64	19	03	-10	63	05	03	-04	69	09	-02	-08	37	06	01	-03	33	07	00	02	92	22
								04	-08	22	100	03	-05	11	66	07	02	22	71	09	-03	14	54	10	03	21	67	-02	05	16	28	08	00	23	53	-03	06	15	39	07	01	16	41	06	-08	18	84
												100	21	07	-02	92	24	09	00	80	24	22	00	91	23	18	00	45	11	04	-20	81	27	23	01	57	19	08	-16	81	30	25	02	83	33	13	05
												21	100	00	-10	22	92	00	00	28	98	04	-05	20	93	04	00	18	43	-07	00	17	63	02	-00	11	52	-03	-02	15	66	-06	01	25	63	-03	-02
												07	00	100	02	12	-01	71	-01	14	00	51	00	12	-01	63	-02	11	-15	-19	12	10	-05	25	-02	03	-15	09	-03	09	-01	-03	-06	03	10	41	03
												-02	-10	02	100	02	-01	10	89	06	-07	09	92	03	00	11	88	-08	05	12	28	08	00	20	63	00	07	19	37	06	00	30	50	10	-13	12	76
																100	25	17	03	90	26	23	02	98	24	24	03	62	16	02	-16	52	32	19	05	08	25	07	-07	81	35	-02	06	83	38	17	07
																25	100	04	07	28	93	08	02	23	99	08	07	22	51	-03	-03	18	68	04	00	16	58	-02	-08	16	69	-08	00	25	61	01	06
																17	04	100	17	13	02	59	10	14	04	94	16	07	-01	36	09	04	04	49	11	02	-05	20	11	03	05	10	10	02	12	67	16
																03	07	17	100	03	02	11	90	04	08	18	100	-10	17	23	34	08	10	22	63	-02	18	17	36	07	10	26	47	10	-06	17	83
																				100	28	23	05	92	27	21	03	65	15	-02	-15	77	33	19	02	71	24	09	-11	76	37	-03	03	75	40	15	08
																				28	100	05	-05	24	94	06	03	23	39	-05	03	21	61	01	02	16	47	-05	-03	19	64	-09	01	30	64	-02	00
																				23	05	100	10	22	07	75	11	11	-04	28	03	15	08	68	06	09	-04	13	-03	16	10	[illegible]	06	15	17	72	14
																				05	-02	10	100	03	03	11	90	-08	11	12	28	11	07	17	57	03	16	17	30	10	07	21	43	14	-06	10	71
																								100	22	21	03	64	13	-01	-15	86	29	18	07	70	23	05	-05	86	33	-04	08	85	17	15	09
																								22	100	07	08	20	49	-03	01	17	63	03	01	14	55	-02	-04	15	65	-08	02	24	56	[illegible]	[illegible]
																								21	07	100	17	11	02	39	11	10	08	57	14	09	-02	15	10	12	09	08	14	10	17	66	17
																								03	08	17	100	-11	19	23	32	08	12	21	62	-02	20	18	35	07	11	25	46	10	-05	17	82
																												100	08	-09	02	61	27	00	-08	82	12	-13	-02	62	27	-20	-05	56	39	08	-03
																												08	100	05	01	14	74	04	09	12	89	03	05	12	67	02	12	12	35	-07	13
																												-09	05	100	05	-13	-01	59	17	-04	03	33	16	-12	-03	34	17	-05	-08	54	21
																												02	01	05	100	73	-04	02	21	-06	-03	-02	45	-14	-03	01	11	-15	00	-05	32
																																100	30	07	14	68	21	-05	00	98	34	-07	15	82	37	05	12
																																30	100	11	05	25	80	-01	-03	28	98	04	10	32	83	01	08
																																07	11	100	26	-05	03	52	18	06	12	59	29	14	11	66	24
																																14	05	26	100	-04	08	23	71	11	06	25	94	11	00	03	57
																																				100	17	-20	-07	75	27	-27	-02	58	35	01	00
																																				17	100	02	01	21	80	00	12	22	48	-08	14
																																				-20	02	100	13	-08	00	71	22	05	-04	33	17
																																				-07	01	13	100	-03	-01	17	76	-08	-04	-01	36
																																								100	32	-12	12	84	36	03	09
																																								32	100	04	11	35	87	01	08
																																								-12	04	100	24	-01	-03	30	23
																																								12	11	24	100	11	04	04	43
																																												100	40	05	09
																																												40	100	08	-02
																																												05	08	100	20
																																												09	-02	20	100

*decimal points have been omitted

Table 6.41

Varimax Rotated Principal Component Factor Analysis of Error of Omission Score

SCORING PROCEDURE	CPMP	FACTOR									h_j^2**
		I	II	III	IV	V	VI	VII	VIII	IX	
AUTHOR	1	87	10	-02	04	01	07	08	07	-10	80
	2	32	-06	09	42	71	20	06	-04	-04	84
	3	14	08	36	11	06	08	22	-18	48	48
	4	-07	64	06	00	01	-02	-06	05	00	43
GCS	1	89	06	-01	11	07	03	00	09	-03	82
	2	26	-09	09	26	90	12	06	-04	-01	98
	3	06	12	83	00	05	-07	-13	20	08	78
	4	09	85	10	-01	03	00	08	07	06	75
GDS	1	85	-01	05	14	14	00	02	03	08	76
	2	15	-14	14	41	74	02	-02	-02	-03	77
	3	01	13	86	-05	02	-07	-05	21	13	83
	4	05	75	15	00	-03	-04	18	00	00	62
ICS	1	88	-04	10	11	03	04	-07	13	04	83
	2	12	-05	-02	89	32	17	02	00	00	95
	3	05	00	62	-01	07	-10	01	-12	-74	97
	4	05	88	03	-03	-10	01	18	11	-03	83
IDS	1	92	-01	15	09	07	09	-02	02	02	90
	2	14	04	02	88	29	26	-06	-03	03	95
	3	02	08	85	02	04	03	11	-07	-23	81
	4	03	94	09	04	-07	11	14	06	04	94
ICM	1	86	02	14	10	11	07	-06	-01	01	78
	2	17	-02	-01	90	32	12	01	-04	01	96
	3	15	06	78	03	08	-02	-01	03	-02	65
	4	07	83	02	-01	-08	10	14	07	-04	74
IDM	1	95	00	12	07	06	06	00	00	-01	93
	2	13	05	01	92	23	23	-03	-04	03	97
	3	10	08	88	04	05	05	15	-13	-12	85
	4	03	92	08	04	-08	14	13	07	04	91
CCS	1	64	-07	07	04	23	-03	-03	-28	08	56
	2	06	09	-05	26	14	83	03	-01	07	81
	3	-07	11	49	-02	-10	08	09	23	57	66
	4	-18	32	04	-01	10	-06	21	-18	-02	23
CDS	1	91	06	00	01	12	06	06	-05	-07	86
	2	21	06	-01	34	60	65	-02	05	03	95
	3	10	11	68	00	04	04	18	46	22	78
	4	07	54	01	01	-03	02	76	16	01	87
CCM	1	72	-02	02	-04	19	04	-04	-34	13	70
	2	15	11	-08	31	21	85	01	02	06	91
	3	00	11	24	-01	-05	02	09	70	-01	57
	4	-09	29	02	-04	06	-04	78	-01	09	72
CDM	1	93	05	00	-01	11	06	05	-10	-05	89
	2	24	05	00	36	64	60	01	06	-01	96
	3	-07	19	18	-07	04	01	07	80	06	74
	4	08	36	02	-11	01	07	84	18	06	89
Mc	1	92	06	06	13	10	02	01	10	-02	88
	2	20	-06	-06	32	85	21	02	-01	-01	96
	3	05	11	89	-02	05	-07	-12	19	16	88
	4	05	93	11	00	03	01	09	03	07	89
Percentage of Variance		20.10	13.75	12.29	8.89	8.37	5.79	4.76	4.11	2.84	80.90

*factor loading ≥ 45; decimal point omitted
**communality

Since only 2.84% of the variance was accounted for by factor IX, it was dropped from further discussions.

Based upon the above factor loadings, the following names were attached to factors I-VIII: CPMP 1 factor; CPMP 4, author, group and individual factor; CPMP 3 group and individual factor; CPMP 2 individual factor; CPMP 2, author, group and computer factor; CPMP 2 computer (single key) factor; CPMP 4 computer (multiple key) factor; and, CPMP 3 computer (multiple key) factor.

The percentage of variance accounted for by each factor is presented at the bottom of Table 6.41. In total, the first eight factors accounted for 78.06 of the observed score variance. Of this, factors I-VIII respectively accounted for 20.10%, 13.75%, 12.29%, 8.89%, 8.37%, 5.79%, 4.76%, and 4.11% of the variance.

The above results suggested a minor clinical problem component (i.e., factor I) and a predominant scoring procedure component (i.e., factors II-VII). Factor I accounted for 20.10% of the total variance and factors II-VIII accounted for 57.96%.

Further factor analytical investigations were undertaken to determine the underlying structure among error of omission scores within each CPMP.

H. Component Analysis of Error of Omission Scores on Each CPMP

The procedure for component analysis of scores for CPMPs 1-4 was the same as that undertaken for proficiency and error of commission scores (see page 129).

a) CPMP 1

Table 6.42 presents the correlation coefficients between error of omission scores on CPMP 1.

Table 6.42

Correlation Coefficient* Between Errors of Omission on CPMP 1 Calculated Using the 12 Scoring Procedures (N = 111)

	AUTHOR	GCS	GDS	ICS	IDS	ICM	IDM	CCS	CDS	CCM	CDM	Mc
AUTHOR	100	87	72	75	79	73	82	45	85	55	86	87
GCS		100	84	81	79	70	81	51	80	55	83	99
GDS			100	77	81	75	87	03	74	63	76	88
ICS				100	92	80	91	45	81	57	81	83
IDS					100	90	48	62	82	08	81	83
ICM						100	92	65	77	71	76	75
IDM							100	64	86	70	86	85
CCS								100	61	82	62	56
CDS									100	68	98	82
CCM										100	75	58
CDM											100	84
Mc												100

*decimal point omitted

Table 6.43 presents the resulting matrix from the principal component analysis.

Table 6.43

Principal Component Factor Analysis of Errors of Omission Scores on CPMP 1

Scoring Procedure	Factor I	h_j^2**
Author	87*	75
GCS	89	79
GDS	87	76
ICS	89	79
IDS	94	89
ICM	88	77
IDM	97	93
CCS	66	44
CDS	92	84
CCM	73	54
CDM	93	86
Mc	92	85
Percentage of Variance	76.86	76.86

* decimal point omitted
**communality

One component was found to underlie the correlation matrix accounting for 76.9% of the observed score variance.

b) CPMP 2

Table 6.44 presents the correlation coefficients between the error of omission scores on CPMP 2.

Table 6.44

Correlation Coefficient* Between Errors of Omission on CPMP 2
Calculated Using the 12 Scoring Procedures (N = 111)

	AUTHOR	GCS	GDS	ICS	IDS	ICM	IDM	CCS	CDS	CCM	CDM	Mc
AUTHOR	100	90	78	68	66	68	63	37	74	51	80	88
GCS		100	84	58	53	58	49	29	74	40	79	97
GDS			100	61	62	61	57	24	61	29	64	82
ICS				100	92	98	93	43	63	52	66	63
IDS					100	93	99	51	68	58	69	61
ICM						100	94	39	61	47	64	64
IDM							100	49	63	55	65	56
CCS								100	74	89	67	35
CDS									100	80	98	83
CCM										100	80	48
CDM											100	43
Mc												100

*decimal point omitted

Table 6.45 presents the resulting matrix from the principal component analysis.

Three components were found to underlie the correlation matrix. Factor loadings were related to the method of categorization of options: author and group

Table 6.45

Varimax Rotated Principal Component Factor Analysis of of Error of Omission Scores on CPMP 2

Scoring Procedure	Factor I	Factor II	Factor III	h_j^2**
Author	79*	39	25	84
GCS	96	22	17	100
GDS	76	39	07	74
ICS	36	87	24	94
IDS	31	87	33	95
ICM	36	89	19	95
IDM	25	91	30	98
CCS	08	22	88	83
CDS	60	29	71	95
CCM	19	26	91	93
CDM	65	31	66	95
Mc	91	28	26	97
Percentage of Variance	34.95	32.18	24.99	91.92

*factor loading ≥ 45; decimal point omitted
**communality

methods loaded on factor I, individual on II, and computer on III. Factors I, II and III respectively accounted for 34.95%, 32.18%, and 24.99% of the total observed variance of 91.92%.

c) CPMP 3

Table 6.46 presents the correlation coefficients between error of omission scores on CPMP 3.

Table 6.46

Correlation Coefficient* Between Errors of Omission on CPMP 3 Calculated Using the 12 Scoring Procedures (N = 111)

	AUTHOR	GCS	GDS	ICS	IDS	ICM	IDM	CCS	CDS	CCM	CDM	Mc
AUTHOR	100	28	24	-13	22	37	40	42	42	-05	-05	31
GCS		100	85	43	63	72	62	41	62	32	31	97
GDS			100	44	68	66	64	63	69	37	33	92
ICS				100	71	51	63	-19	25	09	-03	41
IDS					100	59	94	36	49	20	10	67
ICM						100	75	28	68	13	17	72
IDM							100	39	57	15	08	66
CCS								100	59	33	34	54
CDS									100	52	59	66
CCM										100	71	33
CDM											100	30
Mc												100

*decimal point omitted

Table 6.47 presents the resulting matrix from the principal component analysis.

Table 6.47

Varimax Rotated Principal Component Factor Analysis of Error of Omission Scores on CPMP 3

Scoring Procedure	Factor			h_j^2**
	I	II	III	
Author	18	-11	63*	44
GCS	72	32	35	74
GDS	71	38	40	81
ICS	89	-03	-47	101
IDS	86	07	14	77
ICM	74	11	32	66
IDM	86	01	29	82
CCS	18	35	68	61
CDS	48	55	47	75
CCM	12	79	01	64
CDM	03	88	06	78
Mc	73	33	44	84
Percentage of Variance	39.04	18.41	16.56	74.01

*factor loading ≥ 45; decimal point omitted
**communality

Three components were found to underlie the scoring procedures on CPMP 3. Factor loadings appeared to be related to methods of categorizing options: group and individual methods predominantly loaded on factor I, computer on II, and author on III. Factors I, II and III respectively accounted for 39.04%, 18.41% and 16.56% of the total observed variance of 74.01%.

d) CPMP 4

Table 6.48 presents the correlation coefficients between error of omission scores on CPMP 4.

Table 6.48

Correlation Coefficient* Between Errors of Omission on CPMP 4
Calculated Using the 12 Scoring Procedures (N = 111)

	AUTHOR	GCS	GDS	ICS	IDS	ICM	IDM	CCS	CDS	CCM	CDM	Mc
AUTHOR	100	57	49	56	60	40	59	20	32	12	19	66
GCS		100	78	71	75	67	74	18	55	31	43	97
GDS			100	66	71	54	67	28	53	39	41	84
ICS				100	89	92	88	28	63	37	50	76
IDS					100	90	100	34	63	36	47	83
ICM						100	90	28	57	30	43	71
IDM							100	32	62	35	46	82
CCS								100	21	45	11	32
CDS									100	71	94	57
CCM										100	76	36
CDM											100	43
Mc												100

*decimal point omitted

Table 6.49 presents the resulting matrix from the principal component analysis.

Table 6.49

Varimax Rotated Principal Component Factor Analysis of Error of Omission Scores on CPMP 4

Scoring Procedure	Factor I	II	III	h_j^2**
Author	58*	33	04	45
GCS	83	35	23	86
GDS	74	29	30	72
ICS	45	77	30	88
IDS	54	78	26	97
ICM	32	88	23	93
IDM	51	79	25	95
CCS	19	19	21	12
CDS	26	34	84	89
CCM	14	09	82	70
CDM	14	20	92	91
Mc	89	40	24	100
Percentage of Variance	28.04	27.29	22.86	78.17

*factor loading ≥ 45; decimal point omitted
**communality

Three components were found to underlie the correlation matrix. Factor loadings again appeared to be related to the method of categorizing options: author and group methods loaded predominantly on factor I, individual on II, and computer on III. Factors I, II and III respectively accounted for 28.04%, 27.29%, and 22.86% of the total observed score variance of 78.17%.

I. Discussion of Component Analytic Investigation of Error of Omission scores

A component analysis of the 48 X 48 correlation matrix resulted in a matrix of nine factors. The factor structure yielded the following interpretation:

Factor I: CPMP 1 factor

Factor II: CPMP 4, author, group and individual factor

Factor III: CPMP 3, group and individual factor

Factor IV: CPMP 2, individual factor

Factor V: CPMP 2, author, group and computer factor

Factor VI: CPMP 2, computer (single key) factor

Factor VII: CPMP 4, computer (multiple key) factor

Factor VIII: CPMP 3, computer (multiple key) factor

It was observed that more than one scoring procedure but only one CPMP, loaded on each factor. Therefore, the error of omission score was case specific. However, not all scoring procedures for any given CPMP loaded onto the same factor. Instead, scoring procedures (i.e. more specifically, method of categorization) loaded onto different factors. This variation in structure was more pronounced in the results of the error of omission score analysis. In the analysis of proficiency, error of commission and error of omission scores, approximately 14%, 8% and 60% of the observed score variance was respectively attributed to the effect of the method of categorizing options.

In order to further understand this observed alteration, the scores of each CPMP were factor analyzed. In these investigations, no consistent relationship was observed between the CPMPs and the number of components: in CPMP 2, 3 and 4, three factors were found to underlie the correlation matrices and in CPMP 1, only one factor. A relationship did exist however between methods used to categorize options within scoring procedures and the loadings on each factor, but this relationship was again inconsistent over CPMPs as illustrated below:

CPMP 1:

Factor I: author, group,individual and computer

CPMP 2:

Factor I: author and group
Factor II: individual
Factor III: computer

CPMP 3:

Factor I: group and individual
Factor II: computer
Factor III: author

CPMP 4:

Factor I: author and group
Factor II: individual
Factor III: computer

It was observed that both the simulated clinical problems and categorization methods determined the linear relationship among scores, however, the method of categorization had a greater effect (i.e., accounting for approximately 60% of the observed variance).

J. Component Structure Underlying Efficiency Scores

The correlation coefficients between efficiency scores for the four CPMPs are presented in Table 6.50.

Table 6.51 presents the resulting matrix from the principal component analysis. Eight components were found to underlie the correlation matrix. The following loadings were observed:

CPMP 1, scored by all twelve scoring procedures, loaded on factor I,

CPMP 2, scored by the author, GCS, GDS, ICS, IDS, ICM, IDM, and Mc methods, loaded on factor II,

CPMP 3, scored by the author, GCS, GDS, ICS, IDS, ICM, IDM and Mc methods, loaded on factor III,

CPMP 4, scored by the ICS, ICM, CCS, CDS, CCM and CDM methods, loaded on factor IV,

CPMP 4, scored by the author, GCS, GDS, IDM and Mc methods loaded on factor V,

Table 6.50

Correlation Coefficient* Between Efficiency Scores Calculated Using
12 Scoring Procedures on 4 CPMPs

SCORING PROCEDURE

AUTHOR				GCS				GDS				ICS				IDS				ICM				IDM				CCS				CDS				CCM				CDM				Mc			
C P M P																																															
1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
100	21	20	12	78	20	19	17	78	20	19	17	80	19	16	18	66	10	06	11	88	16	21	23	76	13	14	09	76	23	24	01	70	21	-09	12	73	22	21	22	73	19	13	20	77	19	19	14
21	100	17	-05	34	93	10	-14	34	93	10	-14	27	93	20	-02	28	75	17	-19	18	91	21	-00	32	63	24	-24	24	58	15	-01	25	51	05	02	27	60	11	07	27	57	06	07	34	92	10	-10
20	17	100	22	17	21	81	14	17	21	81	14	18	10	88	22	16	09	57	20	17	11	92	24	15	07	71	16	13	15	69	20	11	12	-18	25	12	13	60	20	12	12	31	23	18	20	81	13
12	-05	22	100	00	-03	21	69	00	-03	21	60	06	-02	17	32	03	-07	17	53	12	-03	18	34	04	-09	24	54	09	04	15	24	07	06	10	51	05	06	18	19	05	01	20	37	01	-06	21	66
				100	34	13	12	100	34	13	12	75	29	18	08	62	24	01	10	74	26	21	09	71	22	08	02	74	23	24	02	69	22	-15	-02	73	24	15	10	73	23	06	07	98	31	13	09
				34	100	15	-10	34	100	15	-10	26	87	23	-04	26	70	18	-17	18	86	23	-03	28	57	26	-21	22	61	17	-06	22	55	-01	00	25	39	11	05	25	59	04	05	34	98	15	-06
				13	15	100	22	13	15	100	22	16	08	69	21	16	05	69	24	16	09	76	22	13	06	71	21	12	12	56	19	10	11	07	27	11	10	44	19	11	13	45	21	12	13	100	22
				12	-10	22	100	12	-10	22	100	10	-12	12	41	08	-11	15	77	17	-12	14	43	09	-05	22	72	15	-01	18	27	15	-01	20	41	12	-04	20	35	11	-04	27	37	12	-11	22	98
								100	34	13	12	75	29	18	08	62	24	01	10	74	26	21	09	71	22	08	02	73	23	24	02	69	22	-15	-02	73	24	15	10	73	23	06	07	98	31	13	09
								34	100	15	-10	26	87	23	-04	26	70	18	-17	18	86	24	-03	28	57	26	-21	22	61	17	-06	22	56	-01	00	25	59	11	05	25	59	04	05	34	98	15	-06
								13	15	100	22	16	08	69	21	16	05	69	24	16	09	76	22	13	06	71	21	12	12	56	19	10	11	07	27	11	10	44	19	11	31	45	21	12	13	100	22
								12	-10	22	100	10	-12	12	41	08	-11	15	77	17	-12	14	43	09	-05	22	72	15	-01	18	27	15	-01	20	41	11	-04	20	35	11	-04	27	37	12	-11	22	98
												100	24	14	14	90	16	10	10	82	21	18	18	91	18	17	04	88	20	20	13	85	16	-08	07	84	21	18	14	84	16	13	13	76	26	16	08
												24	100	15	-02	27	81	12	-18	16	98	15	02	30	66	19	-23	20	56	12	-02	22	48	05	04	24	58	06	09	24	57	01	09	29	86	08	-08
												14	15	100	18	13	18	44	15	12	17	96	25	14	14	65	10	10	22	87	21	08	19	-12	25	08	18	74	21	08	20	27	24	18	22	69	10
												14	-02	18	100	08	-04	22	60	17	-01	24	82	13	-04	24	43	12	08	24	69	11	12	17	71	11	14	17	78	11	12	19	76	09	-03	21	40
																100	18	11	08	74	24	17	16	93	17	16	03	76	22	18	15	80	16	-04	06	80	22	15	13	80	17	09	11	59	27	16	06
																18	100	05	-11	07	83	14	-03	21	86	13	-17	09	59	12	-01	10	54	02	-02	11	60	05	04	11	53	-09	03	22	70	05	-08
																11	05	100	17	07	11	52	18	03	03	87	17	08	04	29	21	05	04	31	24	06	08	20	17	06	04	42	21	00	17	69	16
																08	-11	17	100	15	-17	18	40	07	-06	23	93	13	03	19	42	12	04	24	37	11	02	22	29	11	-01	21	34	10	-16	24	74
																				100	14	17	19	78	11	14	10	79	20	20	12	78	19	-05	07	79	22	16	17	79	18	10	13	75	18	16	15
																				14	100	16	03	27	67	19	-24	18	54	11	-01	20	47	03	04	21	55	04	10	21	54	-01	08	26	85	09	-07
																				17	16	100	28	18	09	69	14	14	20	00	23	13	17	-14	27	14	17	68	24	14	19	37	26	21	21	76	12
																				19	03	28	100	20	-03	27	28	17	11	30	71	17	12	08	85	14	14	21	93	14	12	14	90	08	-03	22	39
																								100	21	12	01	81	22	20	15	85	17	-07	10	84	24	19	18	84	20	10	16	68	27	13	07
																								21	100	11	-13	07	49	08	-03	08	45	-03	-04	08	50	04	03	08	46	-04	01	21	59	06	-02
																								12	11	100	22	15	18	49	30	13	16	23	34	12	18	38	25	12	15	36	30	08	26	71	23
																								01	-13	22	100	08	02	16	32	07	02	25	30	07	-01	22	16	07	-04	22	25	02	-21	21	68
																												100	23	16	11	97	19	-09	17	93	23	13	13	93	19	13	13	76	22	12	11
																												23	100	21	02	22	96	06	09	22	94	17	13	22	91	-02	16	22	63	12	01
																												16	21	100	24	13	19	06	26	13	19	85	29	13	21	38	29	22	16	56	15
																												11	02	24	100	11	02	09	79	09	05	13	71	09	-01	03	79	-01	-07	19	28
																																100	17	-08	06	95	23	11	14	95	19	10	12	70	22	10	12
																																17	100	08	10	18	96	13	14	18	94	-06	15	22	57	11	00
																																-08	08	100	10	-08	09	22	09	-08	08	26	14	-15	02	07	21
																																06	10	10	100	05	10	15	78	05	09	09	93	-04	-01	27	41
																																				100	23	12	11	100	19	14	11	73	25	11	08
																																				23	100	13	17	23	92	-03	17	24	61	10	-04
																																				12	13	100	17	12	15	51	20	15	12	44	17
																																				11	17	17	100	29	71	14	14	09	78	11	17
																																								100	19	14	11	73	25	11	08
																																								19	100	-02	15	23	61	13	-03
																																								14	-02	100	12	06	04	45	24
																																								11	15	12	100	05	04	21	36
																																												100	31	12	09
																																												31	100	13	-08
																																												12	13	100	22
																																												09	-08	32	100

*decimal points have been omitted

Table 6.51

Varimax Rotated Principal Component Factor Analysis of Efficiency Scores

SCORING PROCEDURE	CPMP	Factor								h_j^2**
		I	II	III	IV	V	VI	VII	VIII	
AUTHOR	I	82*	06	09	11	08	08	05	-16	74
	2	19	93	08	01	-11	12	04	06	74
	3	08	08	83	12	04	03	33	-20	87
	4	02	-02	14	21	64	01	03	05	48
GCS	1	82	21	05	-03	09	04	08	-38	88
	2	17	91	13	-03	-07	16	02	-01	91
	3	08	03	95	-08	13	05	06	-01	93
	4	09	-06	09	20	95	-05	04	02	96
GDS	1	82	21	05	-03	09	04	08	-38	88
	2	17	91	13	-03	-07	16	02	-01	91
	3	08	03	95	-08	13	05	06	-01	93
	4	09	-06	09	20	95	-05	04	02	96
ICS	1	92	11	08	07	01	01	03	04	88
	2	15	93	03	04	-09	10	01	06	92
	3	04	13	70	14	01	08	58	-21	92
	4	08	-05	12	76	31	09	03	04	70
IDS	1	84	13	08	07	-02	01	02	14	75
	2	05	80	01	-01	-05	23	03	-07	71
	3	01	11	77	12	28	-06	-13	30	73
	4	08	-16	12	24	79	08	05	07	74
ICM	1	86	03	07	06	09	08	03	-07	78
	2	12	94	04	05	-08	09	-00	03	91
	3	09	11	78	16	02	06	48	-18	92
	4	12	-03	11	91	22	05	08	-01	90
IDM	1	89	16	04	11	-02	01	07	07	84
	2	06	67	01	-03	-00	20	01	-09	50
	3	06	18	78	19	12	02	05	20	74
	4	04	-22	12	13	76	10	06	13	69
CCS	1	93	06	05	05	05	07	-01	05	89
	2	14	49	05	05	03	82	07	02	94
	3	12	07	50	18	07	10	73	-07	85
	4	06	-05	13	79	14	-02	02	06	67
CDS	1	93	06	03	05	04	06	-02	12	89
	2	10	42	04	07	03	90	03	-01	99
	3	-10	04	04	06	22	06	02	49	31
	4	-01	01	17	87	24	02	-00	03	85
CCM	1	94	08	04	03	02	05	-01	11	91
	2	15	49	04	10	-01	82	03	05	95
	3	11	03	37	06	14	06	89	17	99
	4	09	06	08	90	13	04	06	-03	86
CDM	1	94	08	04	03	02	05	-01	11	91
	2	11	48	05	07	-02	80	05	01	88
	3	10	-01	39	-00	21	09	25	30	37
	4	06	05	11	94	18	06	06	03	95
Mc	1	82	21	04	-04	09	04	07	-38	88
	2	16	90	11	-03	-08	19	03	04	89
	3	08	03	95	08	13	05	06	-01	93
	4	06	-01	09	20	92	-06	00	03	90
Percentage of Variance		20.10	15.52	13.55	10.35	10.06	6.52	4.50	2.50	83.10

*factor loading ≥ 45; decimal point omitted

**communality

CPMP 2, scored by the CCS, CDS, CCM and CDM methods, loaded on factor VI,

CPMP 3, scored by the CCS and CCM methods, loaded on factor VII, and

CPMP 3, scored by the CDS method, loaded on factor VIII.

Factor VIII was dropped since the amount of variance it accounted for was too small (i.e., 2.50%).

Based upon the above factor loadings, factors I-VII were referred to by the following names: CPMP 1 factor; CPMP 2, author, group and individual factor; CPMP 3, author, group and individual factor; CPMP 4, individual and computer factor; CPMP 4 author and group factor; CPMP 2 computer factor; and, CPMP 3 computer factor.

The percentage of variance accounted for by each factor is presented at the bottom of Table 6.51. The seven factors respectively accounted for 20.10%, 15.52%, 13.55%, 10.35%, 10.06%, 6.52%, and 4.50% of the total observed score variance of 83.10%.

The percentage of variance accounted for by each factor declined gradually from factor I-VIII. It, therefore, was not possible to identify whether the problem or the scoring procedure had a predominant effect upon the linear relationship of efficiency scores. Further component analysis was undertaken to determine the underlying structure among efficiency scores within each CPMP.

K. Component Analysis of Efficiency Scores on Each CPMP

The procedure for component analysis of scores of

CPMPs 1-4 was the same as that undertaken for proficiency, error of commission and error of omission scores (see page 129).

a) CPMP 1

Table 6.52 presents the correlation coefficients between efficiency scores on CPMP 1.

Table 6.52

Correlation Coefficient* Between Efficiency Scores on CPMP 1
Calculated Using the 12 Scoring Procedures (N = 111)

	AUTHOR	GCS	GDS	ICS	IDS	ICM	IDM	CCS	CDS	CCM	CDM	Mc
AUTHOR	100	78	78	80	66	88	76	76	70	73	73	77
GCS		100	100	75	62	74	71	74	69	73	73	98
GDS			100	75	62	74	71	74	69	73	73	98
ICS				100	90	82	91	88	85	84	84	76
IDS					100	74	93	76	80	80	80	59
ICM						100	78	79	78	79	79	75
IDM							100	81	85	84	84	68
CCS								100	97	93	93	76
CDS									100	95	95	70
CCM										100	100	73
CDM											100	73
Mc												100

*decimal point omitted

Table 6.53 presents the resulting matrix from the principal component analysis.

Table 6.53

Varimax Rotated Principal Component Factor Analysis of Efficiency Scores on CPMP 1

Scoring Procedure	Factor I	Factor II	h_j^{2**}
Author	59	62*	73
GCS	40	91	99
GDS	40	91	99
ICS	81	48	88
IDS	83	31	78
ICM	68	55	76
IDM	83	40	85
CCS	82	46	88
CDS	88	37	92
CCM	86	42	91
CDM	86	42	91
Mc	41	89	96
Percentage of Variance	52.15	36.03	88.18

*factor loading ≥ 45; decimal point omitted
**communality

Two components were found to underlie the correlation matrix. The factors appeared to be related to the method of categorizing options. Factor I would be best referred to as the individual and computer categorization and factor II, as the author and group categorization. Factors I and II respectively accounted for 52.15% and 36.1% of the total observed score variance of 88.18%.

b) CPMP 2

Table 6.54 presents the correlation coefficients between efficiency scores on CPMP 2.

Table 6.54

Correlation Coefficient* Between Efficiency Scores on CPMP 2
Calculated Using the 12 Scoring Procedures (N = 111)

	AUTHOR	GCS	GDS	ICS	IDS	ICM	IDM	CCS	CDS	CCM	CDM	Mc
AUTHOR	100	93	93	93	75	91	63	58	51	60	57	92
GCS		100	100	87	70	86	57	61	55	59	59	98
GDS			100	87	70	86	57	61	56	59	59	98
ICS				100	81	98	66	56	48	58	57	86
IDS					100	83	86	59	54	60	53	70
ICM						100	67	54	47	55	54	85
IDM							100	49	45	50	46	59
CCS								100	96	94	91	63
CDS									100	96	94	57
CCM										100	92	61
CDM											100	61
Mc												100

*decimal point omitted

Table 6.55 presents the resulting matrix from the principal component analysis.

Table 6.55

Varimax Rotated Principal Components Factor Analysis of Efficiency Scores on CPMP 2

Scoring Procedure	Factor		h_j^{2**}
	I	II	
Author	92*	29	93
GCS	89	32	90
GDS	89	32	90
ICS	92	27	92
IDS	75	36	69
ICM	92	25	91
IDM	62	31	48
CCS	35	91	94
CDS	26	97	100
CCM	35	90	94
CDM	34	88	88
Mc	88	34	89
Percentage of Variance	52.27	34.33	86.60

*factor loading ≥ 45; decimal point omitted
**communality

Two components were found to underlie the correlation matrix. Factors were related to the method of categorizing options: the author, group and individual methods loaded on factor I and computer on II. Factors I and II respectively accounted for 52.27% and 34.33% of the total observed variance of 86.60%.

c) CPMP 3

Table 6.56 presents the correlation coefficients between efficiency scores on CPMP 3.

Table 6.56

Correlation Coefficient* Between Efficiency Scores on CPMP 3
Calculated Using the 12 Scoring Procedures (N = 111)

	AUTHOR	GCS	GDS	ICS	IDS	ICM	IDM	CCS	CDS	CCM	CDM	Mc
AUTHOR	100	81	81	88	57	92	71	69	-18	60	31	81
GCS		100	100	69	69	76	51	56	07	44	45	100
GDS			100	69	69	76	71	56	07	44	45	100
ICS				100	44	96	65	87	-12	74	27	69
IDS					100	52	87	29	31	20	42	69
ICM						100	69	80	-14	68	37	76
IDM							100	49	23	38	36	71
CCS								100	06	85	38	56
CDS									100	22	26	07
CCM										100	51	44
CDM											100	45
Mc												100

*decimal point omitted

Table 6.57 presents the resulting matrix from the principal component analysis.

Table 6.57

Varimax Rotated Principal Component Factor Analysis of Efficiency Scores on CPMP 3

Scoring Procedure	Factor I	II	III	h_j^2**
Author	72*	59	-23	92
GCS	90	33	05	92
GDS	90	33	05	92
ICS	51	80	-18	93
IDS	81	06	33	76
ICM	62	73	-18	94
IDM	77	28	22	71
CCS	29	87	07	84
CDS	06	-03	78	61
CCM	11	94	32	100
CDM	33	31	38	35
Mc	90	33	05	92
Percentage of Variance	42.09	30.56	9.52	82.17

*factor loading ≥ 45; decimal point omitted
**communality

Three components were found to underlie the correlation matrix. Factors were dependent upon the method of categorizing options: author, group and individual methods loaded on factor I; author, individual and computer on II; and computer on III. Factors I, II and III respectively accounted for 42.09%, 30.56% and 9.52% of the total observed score variance of 82.17%.

d) CPMP 4

Table 6.58 presents the correlation coefficients between efficiency scores on CPMP 4.

Table 6.58

Correlation Coefficient* Between Efficiency Scores on CPMP 4
Calculated Using the 12 Scoring Procedures (N = 111)

	AUTHOR	GCS	GDS	ICS	IDS	ICM	IDM	CCS	CDS	CCM	CDM	Mc
AUTHOR	100	69	69	32	53	34	54	24	51	19	37	66
GCS		100	100	41	77	43	72	27	41	35	37	98
GDS			100	41	77	43	72	27	41	35	37	98
ICS				100	60	82	43	69	71	78	76	40
IDS					100	40	93	42	37	29	34	74
ICM						100	28	71	85	93	90	39
IDM							100	32	30	16	25	68
CCS								100	79	71	79	28
CDS									100	78	93	41
CCM										100	14	17
CDM											100	36
Mc												100

*decimal point omitted

Table 6.59 presents the resulting matrix from the principal component analysis.

Table 6.59

Varimax Rotated Principal Component Factor Analysis of Efficiency Scores on CPMP 4

Scoring Procedure	Factor		h_j^2**
	I	II	
Author	22	66*	48
GCS	20	96	95
GDS	20	96	95
ICS	78	32	70
IDS	25	82	73
ICM	92	23	90
IDM	14	79	65
CCS	79	17	65
CDS	87	26	83
CCM	92	13	86
CDM	96	18	95
Mc	20	92	88
Percentage of Variance	40.45	39.36	79.81

*factor loading ≥ 45; decimal point omitted
**communality

Two components were found to underlie the correlation matrix. The computer method loaded on factor I while the author and group methods loaded on II. The individual method loaded on both I and II. Factors I and II respectively accounted for 40.56% and 39.36% of the total observed variance of 79.81%.

L. Discussion of the Component Analytic Investigation of Efficiency Scores

A component analysis of the 48 x 48 correlation matrix resulted in seven factors which when interpreted were given the following names:

Factor I: CPMP 1 factor

Factor II: CPMP 2, author, group and individual factor

Factor III: CPMP 3, author, group and individual factor

Factor IV: CPMP 4, individual and computer factor

Factor V: CPMP 4, author and group factor

Factor VI: CPMP 2, computer factor

Factor VII: CPMP 3, computer factor

It was observed that more than one scoring procedure, but only one CPMP, loaded on each factor. Therefore, efficiency, as measured by the computer simulation, was case specific. However, not all scoring procedures loaded on the same factor for any given CPMP. Instead, scoring procedures (i.e., more specifically, method of categorization) loaded on different factors. This variation in structure was similar to that observed in the error of omission results.

In order to further understand this observed alteration, the scores of each CPMP were factor analyzed. In these investigations it was observed that no consistent relationship existed between the CPMP and the number of factors: two factors were observed in CPMPs 1, 2 and 4, and three in CPMP 3. There was however a relationship between the factor loadings and the method of categorizing options but this relationship was not consistent over CPMPs as illustrated below:

CPMP 1:

Factor I: individual and author
Factor II: author and group

CPMP 2:

Factor I: author, group and individual
Factor II: computer

CPMP 3:

Factor I: author, group and individual
Factor II: computer

CPMP 4:

Factor I: computer
Factor II: author, group and individual
Factor III: individual

It was therefore concluded that both CPMP and method of categorization determined the linear relationship among efficiency scores.

M. Summary of the Component Structure Underlying CPMP Scores

Component analyses were undertaken to determine whether the same unitary trait was being assessed. CPMP scores were generated using different scoring procedures. From the analyses of the 48 x 48 correlation matrices, it was observed that only one CPMP and several scoring procedures loaded on each factor. Since the correlations between CPMPs tended to be very small and since no two CPMPs loaded on the same factor, it was concluded that clinical performance, as measured by the computer simulations, was generally case or problem specific. This finding is in keeping with the recent work of Elstein, et al, (1978), who noted that both physicians' and medical students' diagnostic effectiveness varied considerably according to the clinical problem encountered.

There are several possible explanations for the case specificity of physician effectiveness in dealing with clinical problems. This may have occurred because each of the problems had several dimensions interacting (e.g., medical content: obstetrics, cardiology, gynecology, medicine; intervention: history, physical examination, laboratory examination, management; context of care: acute, chronic, health maintenance, emergency; and, structure of simulation: complex, linear branching. If these several dimensions did interact, then calculating a single score (e.g., proficiency) may not have reflected the common elements among cases.

Although the CPMPs were found to be case specific (i.e., only one CPMP fell on each factor), not all scoring procedures fell on the same factor. The scoring procedures were observed to fall on different factors depending on the CPMP, the method of categorization and the score analyzed. For example, when proficiency scores were analyzed, factor VI was referred to as CPMP 2 computer factor, and factor VIII, the CPMP 2 group factor. Since the scoring procedures (more specifically, the methods of categorization) fell on different factors, the scoring procedures may have produced measures of different behaviors.

The 48 x 48 correlation matrices were subdivided and the scores of each CPMP were further analyzed. Through these analyses, it was observed that there was no consistent relationship among the CPMPs and the scoring procedures. The factor loadings were found to be unrelated to:

1) the type of CPMP (branching versus non-branching,
2) the type of weights used within the scoring procedure (constant versus differential), and
3) the type of key used (single versus multiple).

However, a relationship was observed between the factor loadings and the method used to categorize options but this relationship also varied depending on the CPMP and scores analyzed. Table 6.60 summarizes the number of factors found and the groupings of methods used to categorize options.

Table 6.60

Number of Factors and Structure of CPMP Scores

CPMP	Score			
	Proficiency	Error of Commission	Error of Omission	Efficiency
1	1 factor (A,G,I,C)*	1 factor (A,G,I,C)	1 factor (A,G,I,C)	2 factors (I,C) (A,G)
2	3 factors (A,G,) (I) (C)	2 factors (A,G,I) (C)	3 factors (A,G,) (I) (C)	2 factors (A,G,I) (C)
3	2 factors (A,G,I) (C)	3 factors (G,I) (C) (C)	3 factors (A) (G,I) (C)	3 factors (A,G,I) (C) (C)
4	2 factors (A,G,I) (C,I)	2 factors (A,G) (I,C)	3 factors (A,G) (I) (C)	2 factors (C,I) (A,G,I)

* A = author categorizations
G = group categorizations
I = individual categorizations
C = computer categorizations

Table 6.60 illustrates that both the number of factors and the loadings for categorization methods varied over CPMPs and clinical scores. For example, there was one component underlying CPMP 1 and three components underlying the CPMP 3 commission scores. In the error of omission scores calculated on CPMP 2, the individual method loaded by itself on a separate component but loaded with the group method in CPMP 3.

There are two additional observations in Table 6.60 that are worth identifying. Firstly, there was only one component or dimension underlying the proficiency, error of commission and error of omission scores in CPMP 1. Thus, irrespective of the scoring procedure, there was only one component underlying these scores while, in the other CPMPs, there were two or three. This finding may be due to the simplicity of the medical problem simulated in CPMP 1. CPMP 1 was a linearily structured simulation of a 44 year old man with a "straight-forward, easy to diagnose and manage" cardiac problem. Given this conceptually and structurally simple simulated clinical problem, the linear relationship among scores was unaffected by the different methods of categorizing options. Since the other simulated problems were conceptually and structurally more complex, several components were observed to underlie their scores.

Secondly, a pattern of loadings was noted among the methods of categorizing options. The author and group categorization methods loaded on the same factor in 14 out of 16 analyses; the computer method loaded on a component by itself in 9 out of 16 analyses, and the individual method loaded either with the author and group, with the computer or by itself. Thus, there were basically three components underlying the methods of categorization:

1) author and group
2) individual, and
3) computer.

The exact pattern of loadings of these components tended to vary over CPMPs and scores.

Further analyses were carried out to determine the effect that scoring procedures may have upon the means of clinical scores. These results are reported in the next section.

4. Multivariate Analysis

A. Multivariate Analysis of CPMP Scores

Analyses were undertaken to determine the effect that scoring procedures had upon mean scores. The data was subjected to a one-way multivariate analysis with repeated measures.

B. Statistical Analysis

With the enormous amount of data generated within this study (111 examinees X 12 scoring procedures X 4 CPMPs X 3 scores* = 15,984 scores), there were no computer systems or programs available to the author to carry out a multivariate analysis of this data. The computer systems available had insufficient core and the dimensions of available programs were too small. Therefore, a step by step procedure of data analysis was undertaken. The first step in this procedure was to calculate the generalized inverse of large design matrices.

C. Calculation of the Generalized Inverse of a Large Design Matrix

The linear expression of Y, given X, is expressed in Equation 6.1.

$$Y = XB + E \tag{6.1}$$

---

*Although four scores (proficiency, error of commission, error of omission, and efficiency) were calculated for each examinee on each CPMP by twelve different scoring procedures, only the means of three scores (proficiency, error of omission and efficiency) were analyzed. The error of commission was excluded since it was linearly related to the proficiency and error of omission scores (i.e., proficiency + error of commission + error of omission = 100%)

where Y = the observed score

X = the design matrix

$B = (X'X)^{-1} X'Y$, the beta weights, and

E = the error matrix.

Since X'X was singular, X'X inverse could not be found and the pseudo or generalized inverse was used. There are several computer systems and programs in North America that will calculate a generalized inverse if the dimensions of X'X are relatively small, but none if the dimensions are large.

The dimensions of the X'X matrices in this study were large (*i.e.*, 123 X 123). As the generalized inverse solution for this matrix was beyond the scope of the computer systems available, an algebraic solution was sought and found (see Appendix G).

D. One-Way Multivariate Analysis

a) Linear Model

The linear model of the one-way multivariate analysis is expressed in Equation 6.2.

$${}_{1332}Y_{12} = {}_{1332}X_{123}\ {}_{123}B_{12} + {}_{1332}E_{12} \qquad (6.2)$$

where Y = the criterion matrix of 1332 rows and 12 columns
X = the design matrix of 1332 rows and 123 columns
B = the effects matrix of 123 rows and 12 columns, and
E = the error matrix of 1332 rows and 12 columns.

The columns of the design matrix identified the examinee and the scoring procedure used to generate the scores in each row of the criterion matrix. There were 123 columns in the design matrix (111 students + 12 scoring procedures). Since the examinee measurements were repeated on twelve occasions (*i.e.*, 12 scoring procedures), there were 1332 rows in the design matrix.

The columns of the effects matrix represented the same twelve scores as in the criterion matrix. The first 111 rows represented the relative effects that were due to examinees, and the last 12 rows, those that were due to scoring procedures. In total, there were 123 rows in the effects matrix (111 examinees + 12 scoring procedures).

The error matrix was similar in dimension and structure to the criterion matrix and represented the difference between the criterion and the model matrix (*i.e.*, Y - XB).

b) Hypothesis

It was hypothesized that there was no

statistical difference in the mean vector of scores among the twelve scoring procedures. The null and alternate hypotheses were statistically expressed as follows:

$$H_o: \vec{\mu}_{Author} = \vec{\mu}_{GCS} = \vec{\mu}_{GDS} = \vec{\mu}_{ICS} = \vec{\mu}_{IDS} = \vec{\mu}_{ICM} = \vec{\mu}_{IDM} = \vec{\mu}_{CCS} = \vec{\mu}_{CDS} = \vec{\mu}_{CCM} = \vec{\mu}_{CDM} = \vec{\mu}_{Mc}$$

$$H_1: \vec{\mu}_i \neq \vec{\mu}_j$$

where $\vec{\mu}$ = vector of twelve scores in the population
i,j= scoring procedure and $i \neq j$

c) Results

The sum of squares and cross-products due to total (Y'Y matrix), model (B'X'Y), model corrected for sum of squares due to means ($B'X'Y - N\bar{Y}^2$), scoring procedure ($(K'B)'(K'(X'X)^- K)^{-1}(K'B)$), and beta weights ($(X'X)^- X'Y$), and error ($Y'Y - \hat{Y}'\hat{Y}$), are respectively presented as Appendices H - M.

The tests for the one-way multivariate analysis of differences in vectors of mean scores among scoring procedures showed a significant difference (see Table 6.61). The null hypothesis was therefore rejected and the alternate hypothesis accepted.

Table 6.61

Raos Approximate F Test Using Wilks Lambda

$df_1 = 132$, $df_2 = 10{,}731.5$, $F = 252.38$, $P < 0.001$
and Lambda = 0.00001

Ray's Maximum Eigenvalue Test

$s = 11$, $m = 0.0$, $n = 653.5$, Heck = 0.9729
Critical Heck ($\alpha = 0.05$) 0.0406, ($\alpha = 0.01$) 0.0450

d) Simultaneous-Paired Comparisons

Additional analyses were then carried out to determine on which of the 12 CPMP scores (i.e., proficiency, error of omission and efficiency, calculated for each of the four CPMPs) the scoring procedures differed. For each variable, a simultaneous-paired comparison, Morrison, (1967), was made on all pairs of scoring procedure means. There were 66 paired comparisons (i.e., $(n^2 - n)/2$) within each scoring procedure and 792 comparisons in total (i.e., 66 X 4 CPMPs X 3 scores). It was hypothesized that there was no statistical difference among mean scores calculated using different scoring procedures. The null and alternative hypotheses were statistically expressed as follows:

$H_o: \mu_{im} - \mu_{jm} = 0$

$H_1: \mu_{im} - \mu_{jm} \neq 0$

where μ = population mean score

i,j = scoring procedure with $i \neq j$

m = CPMP score (CPMP 1, proficiency; CPMP 1, error of commission, ...CPMP 4, efficiency).

Since it was difficult to interpret all 66 paired comparisons among the twelve scoring procedures, Table 6.62 and subsequent tables of paired comparisons were collapsed to determine whether the following characteristics of the scoring procedures systematically altered CPMP scores:

1) method of categorization, (i.e., Author vs GDS vs IDS vs CDS vs Mc, and GCS vs ICS vs CCS),

2) constant versus differential weights, (i.e., GCS vs GDS, ICS vs IDS, and CCS vs CDS), and

3) single versus multiple keys (i.e., ICS vs ICM, IDS vs IDM, CCS vs CCM and CDS vs CDM).

In the above comparisons the following strategy was employed:

i) method of categorization

This inquiry was divided into two sets of comparisons:

set 1: those scoring procedures employing differential weights and a single key,

and

set 2: those procedures with constant weights and single keys.

In set 1, the author, group, individual,computer and McLaughlin methods were compared, and in set 2, the group, individual and computer methods were compared. With five scoring procedures in set 1, ten pairs of comparisons were made and with three scoring procedures in set 2, three pairs of comparisons were made. Only the method of categorization differed among scoring procedures in sets 1 and 2.

ii) constant versus differential weights:

In determining whether constant and differential weights systematically altered mean scores, five paired comparisons were made: GCS vs GDS, ICS vs IDS, CCS vs CDS, ICM vs IDM, and CCM vs CDM. In each paired comparison only the type of weights differed (i.e., constant vs. differential).

iii) single versus multiple keys:

In determining whether constant and differential weights systematically altered mean scores, four paired comparisons were made: ICS vs ICM, IDS vs IDM, CCS vs CCM, and CDS vs CDM. In each paired comparison only the type of key differed (i.e., single vs. multiple).

The above strategy was employed on the mean proficiency, error of omission, and efficiency scores for each of the four CPMPs giving a total of twelve groups of comparisons.

In the following tables of paired comparisons, statistically equal means are identified by a common underline. Means excluded from the sequence of ordered means are indicated by a dotted line.

d.1) CPMP 1, Proficiency Scores

Table 6.62 presents the multiple comparisons of mean proficiency scores on CPMP 1 calculated using the twelve scoring procedures. Table 6.63 summarizes these results by:

i) method of categorization

set 1:

The author, group and McLaughlin methods were found to be statistically equal and different from the individual and computer methods (see line 1 of Table 6.63). The group, McLaughlin and individual methods were statistically equal and different from the author and computer methods (see line 2 of Table 6.63). The McLaughlin, individual and computer methods were statistically equal and different from the author and group methods (see line 3 of Table 6.63).

Table 6.62

Multiple Comparison of Mean Proficiency Scores on CPMP 1
Calculated Using the 12 Scoring Procedures

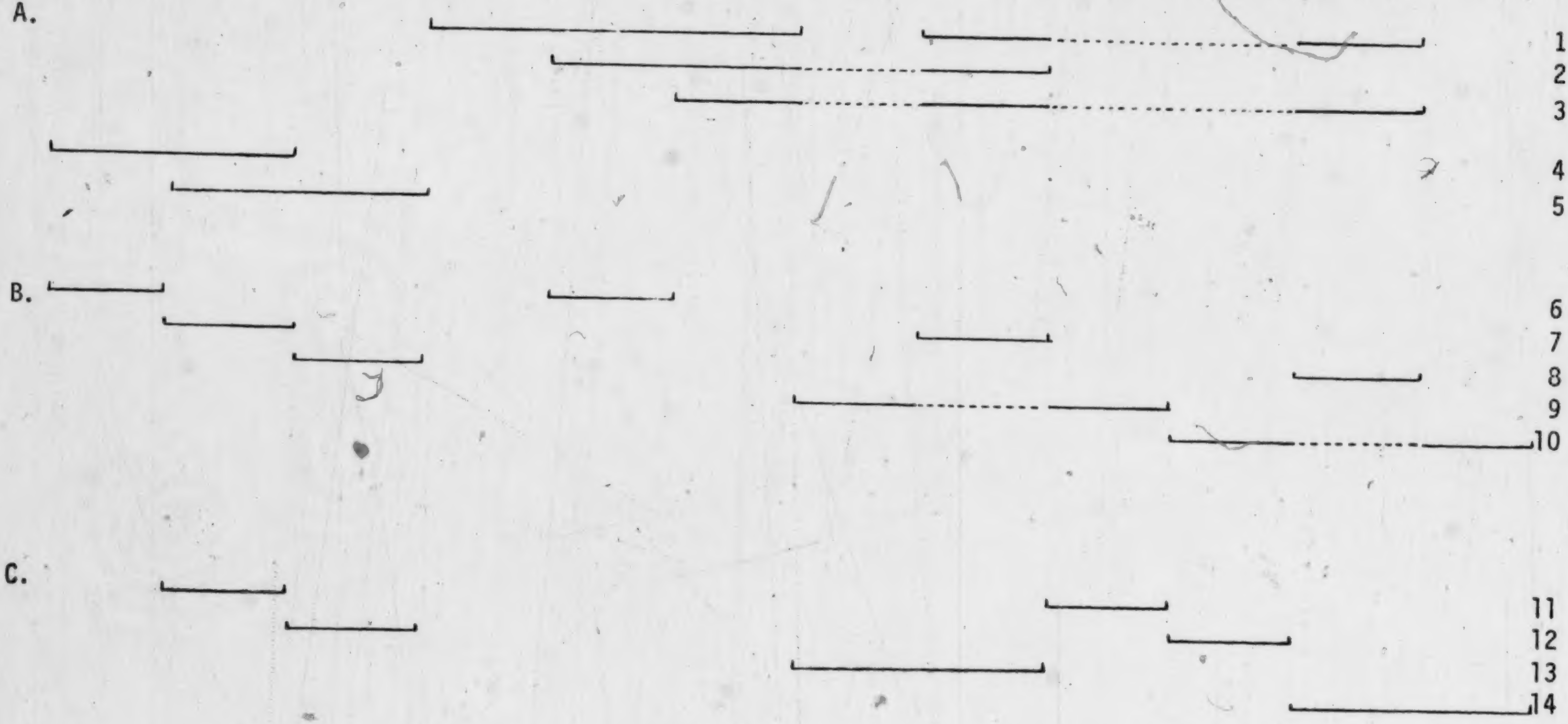
	Author	GCS	GDS	ICS	IDS	ICM	IDM	CCS	CDS	CCM	CDM	Mc
	85.357[a]	81.931	85.673	83.234	87.488	87.856	86.423	84.347	89.131	88.167	89.498	86.401
Author 85.351		3.426**	-0.316	2.123**	-2.131**	-2.499**	-1.066	1.010	-3.774**	-2.810**	4.141**	-1.044
GCS 81.931			-3.742**	1.303	-5.557**	-5.925**	-4.492**	-2.416**	-7.200**	-6.236**	-7.567**	-4.470**
GDS 85.673				2.439**	-1.815	-2.183**	-0.750	1.326	-3.458**	-2.494**	-3.825**	-0.729
ICS 83.234					-4.254**	-4.622**	-3.189**	-1.113	-5.897**	-4.933**	-6.264**	-3.167**
IDS 87.488						-0.368	1.065	3.141**	-1.643	-0.679	-2.010*	1.087
ICM 87.856							1.433	3.509**	-1.275	-0,311	-1.642	1.455
IDM 86.423								2.076**	-2.708**	-1.744	-3.075**	0.022
CCS 84.347									-4.784**	-3.920**	-5.151**	-2.054*
CDS 89.131										-0.964	-0.367	2.730**
CCM 88.167											-1.331	1.766
CDM 89.498												3.097**
Mc 86.401												

[a] mean score for scoring procedure
* 95.0% confidence interval = ± 1.954
**99.0% confidence interval = ± 2.062

Table 6.63

Multiple Comparison of Mean Proficiency Scores on CPMP 1 By:
A. Method of Categorization B. Constant Versus Differential Weights C. Single Versus Multiple Keys

GCS	ICS	CCS	Author	GDS	Mc	IDM	IDS	ICM	CCM	CDS	CDM
81.931	83.234	84.347	85.357	85.673	86.401	86.423	87.488	87.856	88.167	89.131	89.498



set 2:

The group method differed significantly from the computer. The individual method was statistically equal to both the computer and individual methods (see lines 4 and 5 of Table 6.63).

ii) constant versus differential weights

A statistical difference was observed in scoring procedures with single keys: GDS > GCS, IDS > ICS, CDS > CCS; but, no differences were found in scoring procedures with multiple keys: ICM = IDM and CCM = CDM (see lines 6-10 of Table 6.63).

iii) single versus multiple keys

A statistical difference was observed in scoring procedures with constant weights: ICM > ICS and CCM > CCS; but no difference was observed in scoring procedures with differential weights: IDS = IDM and CDS = CDM.

d.2) CPMP 1, Error of Omission Scores

Table 6.64 presents the multiple comparisons of mean error of omission scores on CPMP 1. Table 6.65 summarizes these results by:

i) method of categorization

set 1:

The computer mean error of omission score was found to be statistically lower than those of the author individual, McLaughlin and group methods (see line 1 of Table 6.65).

Table 6.64

Multiple Comparison of Mean Errors of Omission Scores on CPMP 1 Calculated Using the 12 Scoring Procedures

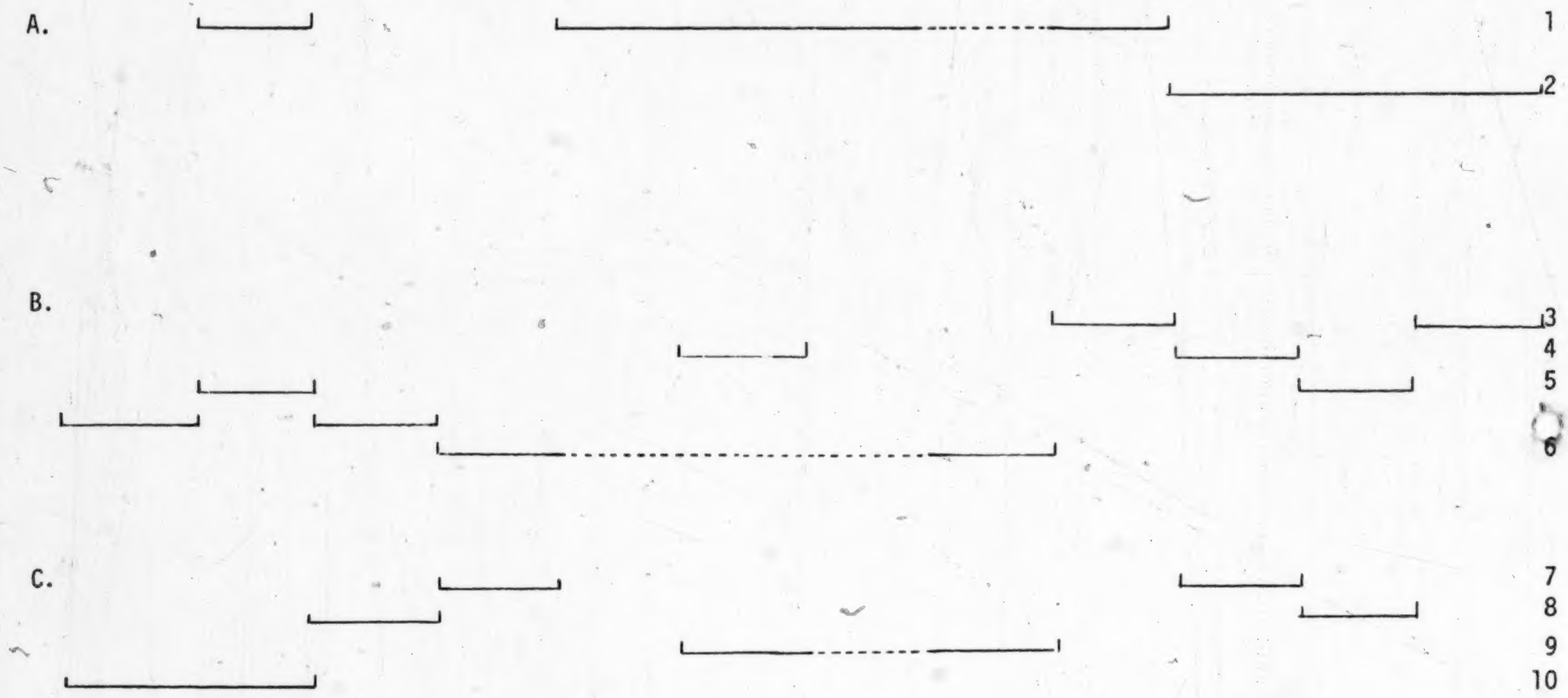
	Author	GCS	GDS	ICS	IDS	ICM	IDM	CCS	CDS	CCM	CDM	Mc
	10.317[a]	14.578	11.524	14.110	10.556	10.321	11.218	14.339	5.239	10.096	5.042	10.591
Author 10.317		-4.261**	-1.207	-3.793**	-0.239	-0.004	-0.901	-4.022**	5.078**	0.221	5.275**	-0.274
GCS 14.579			3.054**	3.468**	4.022**	4.257**	3.360**	0.238	9.339**	4.482**	9.536**	3.987**
GDS 11.324				-2.586**	0.968	1.203	0.306	-2.815**	6.285**	1.428**	6.482**	0.933
ICS 14.110					3.554**	3.789**	2.892**	-0.229	8.871**	4.014**	9.068**	3.519**
IDS 10.556						0.234	-0.661	-3.783**	5.317**	0.460	5.514**	-0.034
ICM 10.321							-0.897	-4.018**	5.082**	0.225	5.279**	-0.269
IDM 11.218								-3.121**	5.979**	1.222	6.176**	0.627
CCS 14.339									9.100**	4.243**	9.297**	3.748**
CDS 5.239										-4.857**	0.197	-5.352**
CCM 10.096											5.050**	0.495
CDM 5.0416												-5.545**
Mc 10.591												

[a] mean score for scoring procedure
* 95.0% confidence interval = ± 1.368
**99.0% confidence interval = ± 2.062

Table 6.65

Multiple Comparison of Mean Error of Omission Scores on CPMP 1 by:
A. Method of Categorization B. Constant Versus Differential Weights C. Single Versus Multiple Keys

CDM	CDS	CCM	ICM	Author	IDS	Mc	IDM	GDS	ICS	CCS	GCS
5.042	5.239	10.096	10.321	10.317	10.556	10.591	11.218	11.524	14.110	14.339	14.578



set 2:

The means of the individual, computer and group consensus methods were statistically equal (see line 2 of Table 6.65).

ii) constant versus differential weights

A statistical difference was observed between five out of six paired comparisons: GCS > GDS, ICS > IDS, CCS > CDS, CCM > CDM; but no difference was found between ICM and IDM.

iii) single versus differential keys

A statistical difference was observed in scoring procedures with constant weights: ICS > ICM and CCS > CCM; but no difference was observed in scoring procedures with differential weights: IDS = IDM and CDS = CDM.

d.3) CPMP 1, Efficiency Scores

Table 6.66 presents the multiple comparisons of mean efficiency scores on CPMP 1. Table 6.67 summarizes these results by:

i) method of categorization

set 1:

The computer method was found to be statistically lower than the author, McLaughlin, individual and group methods (see line 1 of Table 6.67).

set 2:

The computer method was statistically lower than the individual and group methods (see line 2 of Table 6.67).

Table 6.66

Multiple Comparison of Mean Efficiency Scores on CPMP 1
Calculated Using the 12 Scoring Procedures

	Author	GCS	GDS	ICS	IDS	ICM	IDM	CCS	CDS	CCM	CDM	Mc
	84.799[a]	87.969	87.969	87.896	87.300	70.785	86.527	83.636	81.481	77.562	77.562	87.276
Author 84.799		-3.170**	-3.170**	-3.096**	2.500	14.016**	-1.727	1.163	3.318**	7.237**	7.237**	2.476
GCS 87.969			0.000	0.072	0.668	17.186**	1.442	4.332**	6.487**	10.407**	10.407**	-0.693
GDS 87.969				0.072	0.668	17.186**	1.442	4.332**	6.487**	10.407**	10.407**	-0.693
ICS 87.896					0.597	17.113**	1.369	4.259**	6.415**	10.334**	10.334**	-0.620
IDS 87.300						16.517**	0.773	3.663**	5.819**	9.737**	9.737**	-0.024
ICM 70.783							-15.744**	-12.853**	10.698**	-6.779**	-6.779**	-16.492**
IDM 86.527								2.891**	5.045**	8.964**	8.964**	0.748
CCS 83.636									2.155	6.074**	6.074**	3.639**
CDS 81.481										3.918**	3.918**	5.794**
CCM 77.562											0.000	9.714**
CDM 77.562												-9.714**
Mc 87.276												

[a] mean score for scoring procedure
* 95.0% confidence interval = ± 2.631
**99.0% confidence interval = ± 2.775

Table 6.67

Multiple Comparison of Mean Efficiency Scores on CPMP 1 by:
A. Method of Categorization B. Constant Versus Differential Weights C. Single Versus Multiple Keys

	ICM	CCM	CDM	CDS	CCS	Author	IDM	Mc	IDS	ICS	GCS	GDS	
	70.785	77.562	77.562	81.481	83.636	84.799	86.527	87.276	87.300	87.896	87.969	87.969	
A.				—		—	—	—	—	—	—	—	1
					—					—	—		2
B.	—						—						3
											—	—	4
									—	—			5
				—	—								6
		—	—										7
C.		—			—								8
	—									—			9
		—	—										10
							—	—	—				11

ii) constant versus differential weights

Only one significant difference was observed: IDM > ICM. No differences were observed in the other four paired comparisons: GCS = GDS, ICS = IDS, CCS = CDS and CCM = CDM (see lines 3-7 of Table 6.67).

iii) single versus multiple keys

The mean efficiency scores were found to be higher in the following comparisons: CCS > CCM, ICS > ICM and CDS > CDM; but no difference was found between IDM and IDS (see lines 8 - 11 of Table 6.67).

d.4) CPMP 2, Proficiency Scores

Table 6.68 presents the multiple comparisons of mean proficiency scores on CPMP 2. Table 6.69 summarizes these results by:

i) method of categorization

set 1:

The means of author and individual methods were found to be equal but different from the group, McLaughlin and computer methods. The group and McLaughlin methods were equal and differed from the computer, author and individual methods. The computer method was significantly higher than the others. Lastly, the individual, group and McLaughlin means were equal and they differed from the author and computer methods (see lines 1 and 2 of Table 6.69).

set 2:

The means of individual and group methods were equal and they differed significantly from the

Table 6.68

Multiple Comparison of Mean Proficiency Scores on CPMP 2 Calculated Using the 12 Scoring Procedures

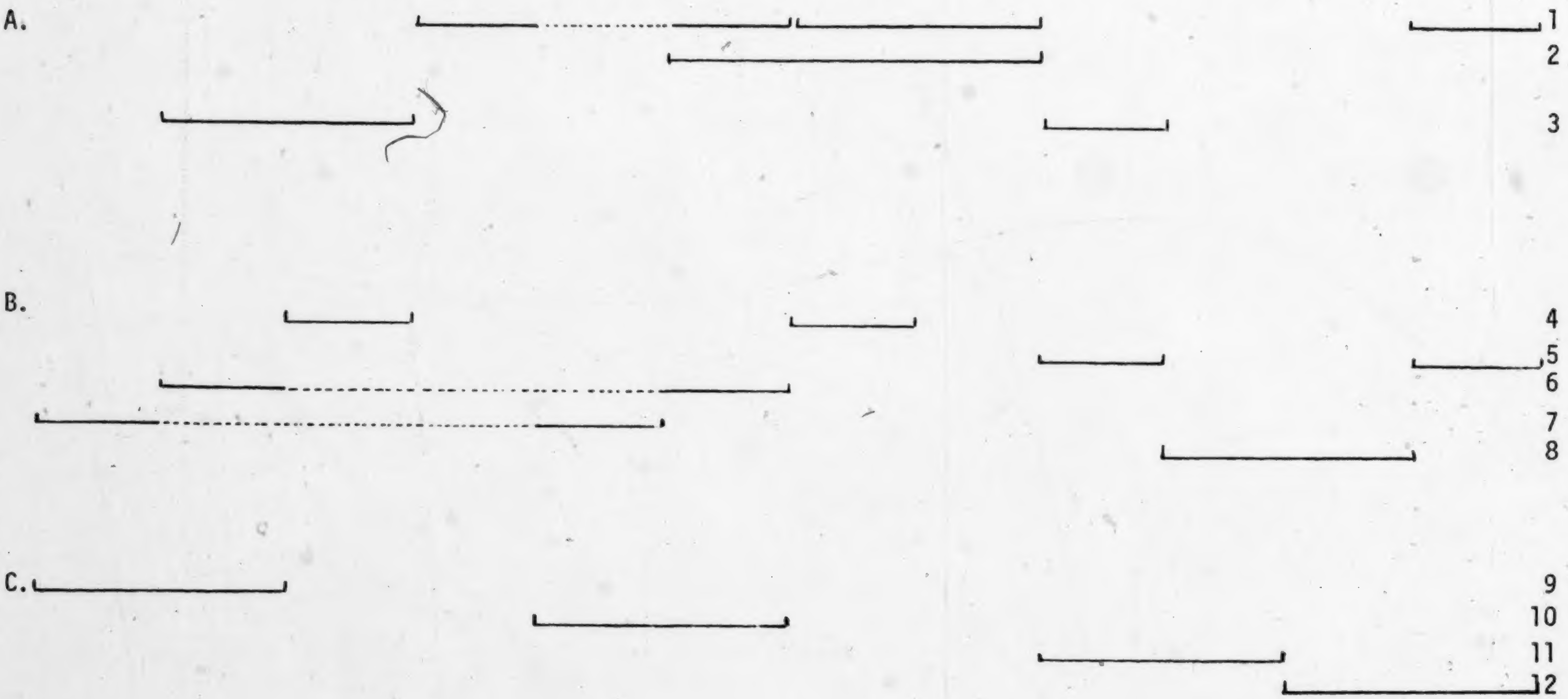
	Author	GCS	GDS	ICS	IDS	ICM	IDM	CCS	CDS	CCM	CDM	Mc
	81.434[a]	80.814	84.512	80.568	82.700	79.738	82.057	86.439	89.582	86.670	89.281	85.084
Author 81.434		0.620	-3.078**	0.865	-1.266	1.695	-0.623	-5.005**	-8.148**	-5.235**	-7.847**	-3.650**
GCS 80.814			-3.697**	0.246	-1.886	1.076	-1.242	-5.625**	-8.768**	-5.855**	-8.467**	-4.270**
GDS 84.512				3.944**	1.811	4.773**	2.454	-1.927	-5.070**	-2.157	-4.769**	-0.572
ICS 80.568					-2.132	0.829	-1.489	-5.871**	-9.014**	-6.101**	-8.713**	-4.516**
IDS 82.700						2.962*	0.643	-3.739**	-6.882**	-3.969**	-6.581**	-2.384
ICM 79.738							-2.319	-6.701**	-9.844**	-6.931**	-9.543**	-5.346**
IDM 82.057								-4.382**	-7.525**	-4.612**	-7.224**	-3.027**
CCS 86.439									-3.143	-0.230	-2.842**	1.355
CDS 89.528										2.912*	0.300	4.498**
CCM 86.670											-2.611	1.585
CDM 89.281												4.197**
Mc 25.084												

[a] mean score for scoring procedure
* 95.0% confidence interval = ± 2.765
**99.0% confidence interval = ± 2.917

Table 6.69

Multiple Comparison of Mean Proficiency Scores on CPMP 2 by:
A. Method of Categorization B. Constant Versus Differential Weights C. Single Versus Multiple Keys

ICM	ICS	GCS	Author	IDM	IDS	GDS	Mc	CCS	CCM	CDM	CDS
79.738	80.568	80.814	81.434	82.057	82.700	84.512	85.084	86.439	86.670	89.281	89.582



computer (see line 3 of Table 6.69).

ii) constant versus differential weights

Two of the five paired comparisons showed a significant difference: GDS > GCS and CDS > CCS. No differences were found in three comparisons: ICS = IDS, ICM = IDM, and CCM = CDM (see lines 4-8 of Table 6.69).

iii) single versus multiple keys

No differences in mean scores were observed between scoring procedures with single and multiple keys: ICS = ICM, IDS = IDM, CCS = CCM, and CDS = CDM (see lines 9-12 of Table 6.69).

d.5) CPMP 2, Error of Omission Scores

Table 6.70 presents the multiple comparisons of mean error of omission scores on CPMP 2. Table 6.71 summarizes these results by:

i) method of categorization

set 1:

The mean scores of the computer and individual methods were found to be different from all others. The means of the McLaughlin, author and group consensus methods were found to be equal (see line 1 of Table 6.71).

set 2:

The means of the group, computer and individual methods were found to be statistically different from each other (see line 2 of Table 6.71).

ii) constant versus differential weights

A difference was noted in mean

Table 6.70

Multiple Comparison of Mean Errors of Commission Scores on CPMP 2
Calculated Using the 12 Scoring Procedures

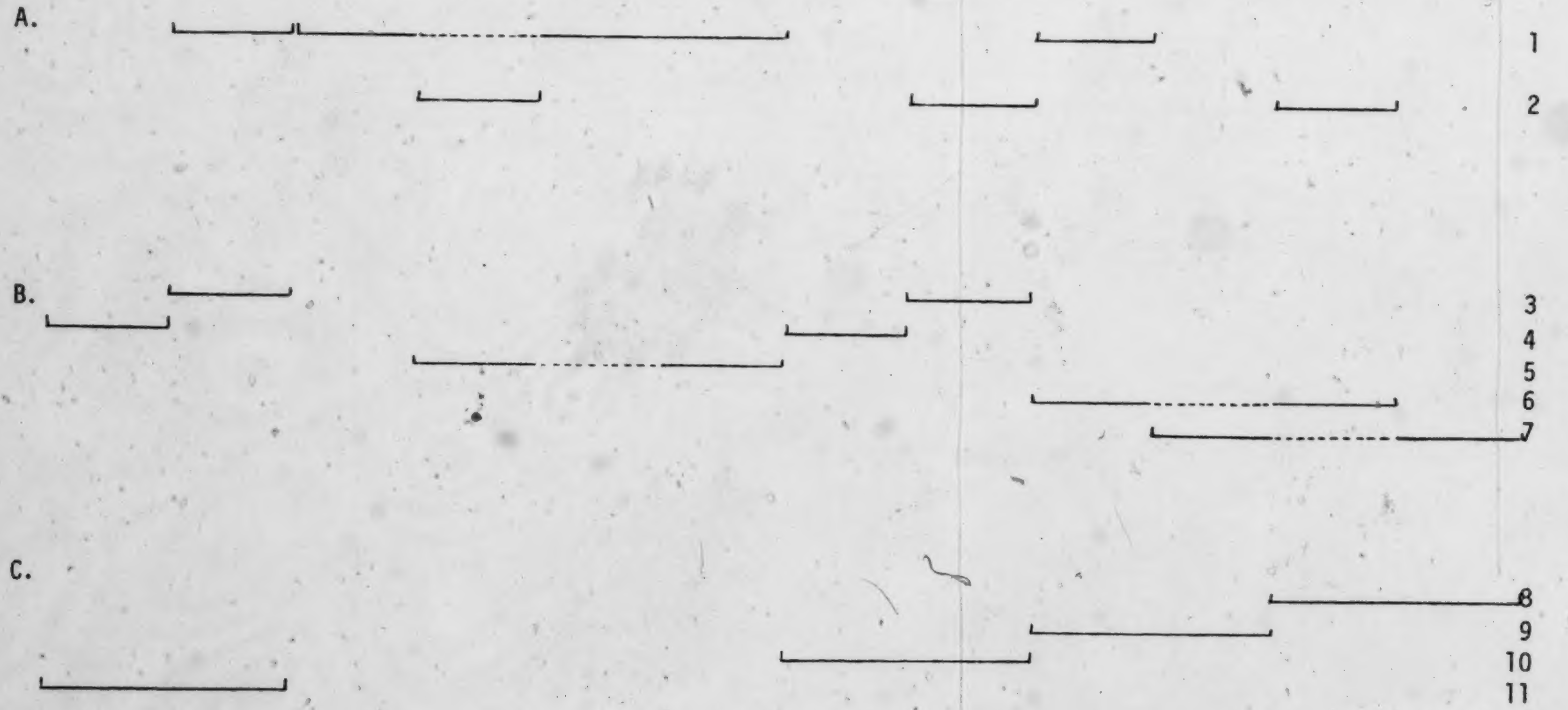
	Author	GCS	GDS	ICS	IDS	ICM	IDM	CCS	CDS	CCM	CDM	Mc
	10.047[a]	9.479	10.082	16.058	14.863	16.981	15.393	12.723	5.108	11.788	4.738	8.945
Author 10.047		0.567	-0.035	-6.011**	-4.816**	-6.934**	-5.346**	-2.676**	4.939**	-1.741	5.309**	-1.102
GCS 9.479			-0.603	-6.579**	-5.384**	-7.502**	-5.914**	-3.244**	4.371**	-2.309**	4.741**	0.533
GDS 10.082				-5.976**	-4.781**	-6.899**	-5.311**	-2.641**	4.974**	-1.706	5.344**	-1.137
ICS 16.058					1.194	-0.923	0.664	3.335**	10.950**	4.270**	11.320**	-7.112**
IDS 14.863						-2.118**	0.529	2.140**	9.755**	3.075**	10.125**	-5.918**
ICM 16.981							1.588	4.258**	11.873**	5.193**	12.243**	-8.036**
IDM 15.393								2.670**	10.285**	3.605**	10.655**	-6.448**
CCS 12.273									7.615**	-0.935	7.985**	-3.778**
CDS 5.108										-6.680**	0.369	-3.837**
CCM 11.788											7.049**	-2.843**
CDM 4.738												-4.207**
Mc 8.945												

[a] mean score for scoring procedure
* 95.0% confidence interval = ± 1.768
**99.0% confidence interval = ± 1.865

Table 6.71

Multiple Comparison of Mean Error of Omission Scores on CPMP 2 by:
A. Method of Categorization B. Constant Versus Differential Weights C. Single Versus Multiple Keys

CDM	CDS	Mc	GCS	Author	GDS	CCM	CCS	IDS	IDM	ICS	ICM
4.738	5.108	8.945	9.479	10.047	10.082	11.788	12.723	14.863	15.393	16.058	16.981



scores of scoring procedures with the computer method: CCS > CDS and CCM > CDM. No statistical differences were observed among the three other paired comparisons: GCS = GDS, ICS = IDS and ICM = IDM (see lines 3-7 on Table 6.71).

iii) single versus multiple keys

No statistical differences were observed between scoring procedures with single and multiple keys: ICS = ICM, IDS = IDM, CCS = CCM and CDS = CDM (see lines 8-11 on Table 6.71).

d.6 CPMP 2, Efficiency Scores

Table 6.72 presents the multiple comparisons of mean efficiency scores on CPMP 2. Table 6.73 summarizes these results by:

i) method of categorization

set 1:

The means of scoring procedures McLaughlin, group and computer were statistically equal while both the author and individual methods produced mean scores that were different from the others (see line 1 of Table 6.73).

set 2:

The mean scores of the group, computer and individual methods were statistically different (see line 2 of Table 6.73).

ii) constant versus differential weights

Four of the five paired comparisons indicated mean scores to be different: IDS > ICS, IDM > ICM,

Table 6.72

Multiple Comparison of Mean Efficiency Scores on CPMP 2 Calculated Using the 12 Scoring Procedures

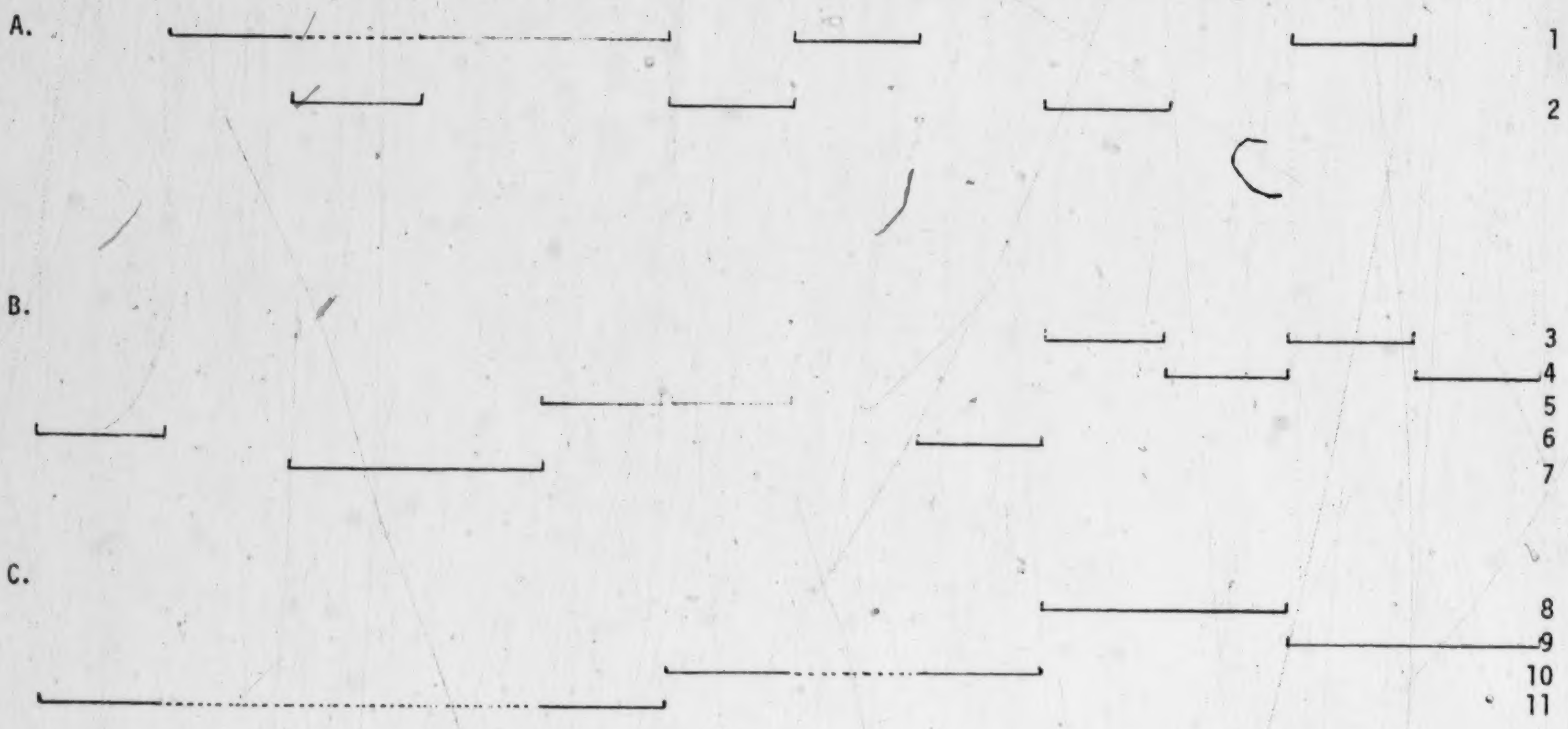
	Author	GCS	GDS	ICS	IDS	ICM	IDM	CCS	CDS	CCM	CDM	Mc
	78.477	73.154	73.154	81.424	88.937	82.011	91.565	77.943	73.671	78.916	70.494	72.884
Author 78.477		5.322**	5.323**	-2.947	-10.454**	-3.534	-13.088**	0.534	4.806**	-0.438	7.983**	-5.593**
GGS 73.154			0.000	-8.270**	-15.777**	-8.857**	-18.411**	-4.788**	-0.516	-5.761**	2.660	-0.270
GDS 73.154				-9.270**	-15.777**	-8.857**	-18.411**	-4.788**	-0.516	-5.761**	2.660	-0.270
ICS 81.424					-7.507**	-0.586	-10.141**	3.481*	7.753**	2.508	10.930	-8.540**
IDS 88.931						6.920**	-2.634	10.988**	15.260**	10.015**	18.437**	-16.047**
ICM 82.011							-9.554**	4.068**	8.340**	3.095	11.517	-9.127**
IDM 91.565								13.622**	17.894**	12.649**	21.071**	-18.681**
CCS 77.943									4.272**	-0.972	7.449**	-5.059**
CDS 93.671										-5.244**	3.177	-8.422**
CCM 78.916											8.422**	-6.032**
CDM 70.494												-2.390
Mc 72.884												

a mean score for scoring procedure
* 95.0% confidence interval = ± 3.377
**99.0% confidence interval = ± 3.562

Table 6.73

Multiple Comparison of Mean Efficiency Scores on CPMP 2 by:
A. Method of Categorization B. Constant Versus Differential Weights C. Single Versus Multiple Keys

CDM	Mc	GCS	GDS	CDS	CCS	Author	CCM	ICS	ICM	IDS	IDM
70.494	72.884	73.154	73.154	73.671	77.943	78.477	78.916	81.424	82.011	88.937	91.565



CCS > CDS and CCM > CDM: but no difference was found between GCS and GDS (see lines 3-7 of Table 6.73).

iii) single versus multiple keys

No statistical differences were found among mean scores of single and multiple keys: ICS = ICM, IDS = IDM, CCS = CCM and CDS = CDM (see lines 8-11 of Table 6.73).

d.7) CPMP 3, Proficiency Scores

Multiple comparisons of mean proficiency scores on CPMP 3 are presented in Table 6.74. There are few statistical differences among scoring procedures (i.e., six out of 66 paired comparisons). The small number of observed differences were primarily due to the large error term which is reflected in the 95% and 99% confidence intervals of ±10.910 and ±11.510. There were no differences in mean scores due to method of categorization, constant and differential weights, and single and multiple keys. These results are summarized in Table 6.75.

d.8 CPMP 3, Error of Omission Scores

Table 6.76 presents the multiple comparisons of mean error of omission scores on CPMP 3. Table 6.77 summarizes these results by:

i) method of categorization

set 1:

The mean score of the individual method was significantly different from the others. The mean scores

Table 6.74

Multiple Comparison of Mean Proficiency Scores on CPMP 3
Calculated Using the 12 Scoring Procedures

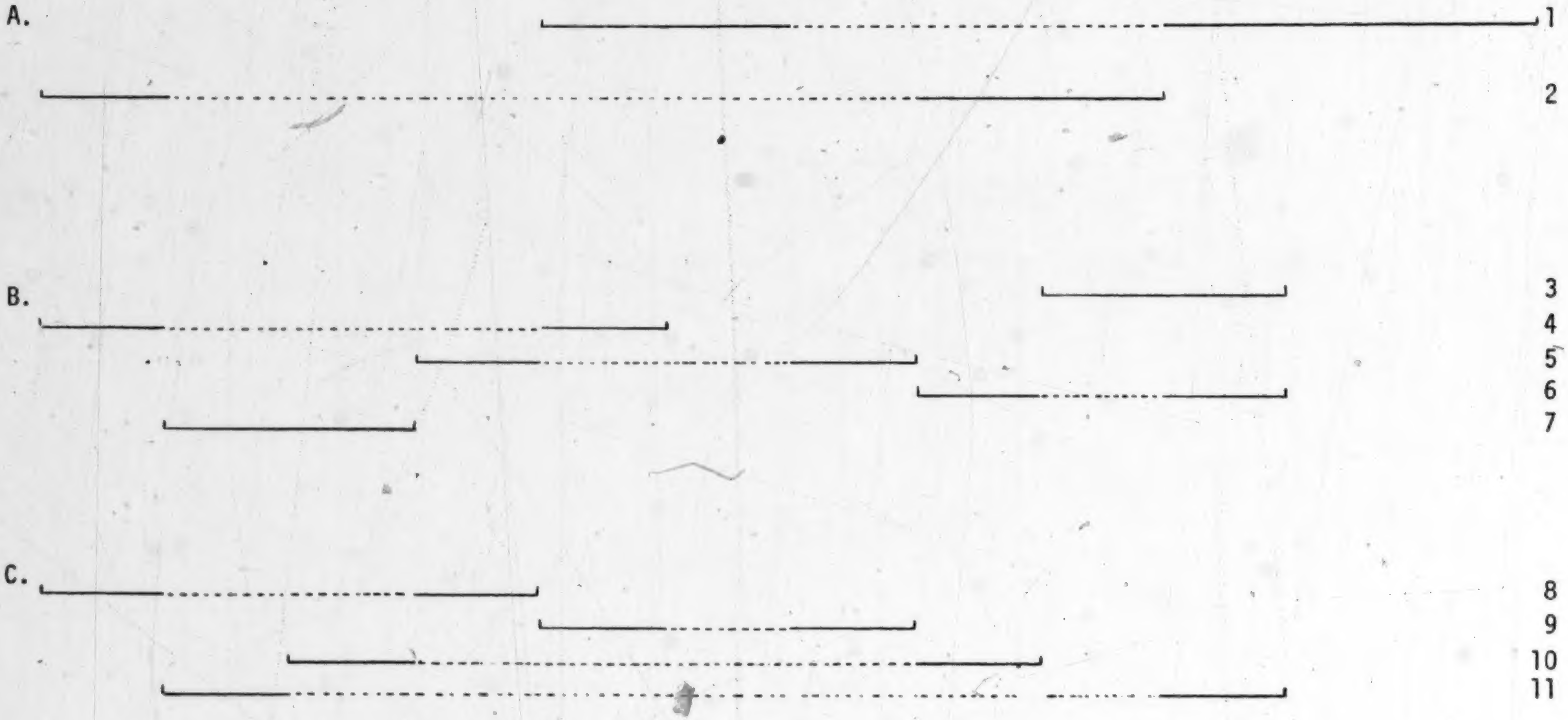
	Author	GCS	GDS	ICS	IDS	ICM	IDM	CCS	CDS	CCM	CDM	Mc
	58.054[a]	64.291	66.978	53.554	57.487	55.015	58.067	60.075	64.372	53.965	53.829	64.641
Author 58.004		-6.237	-8.923	4.500	0.567	3.039	-0.012	-2.021	-6.317	4.088	4.224	-6.592
GCS 64.291			-2.688	10.737	6.803	9.275	6.223	4.215	-0.081	10.326	10.462	-0.356
GDS 66.978				13.424**	9.490	11.963**	8.910	6.902	2.605	13.012**	13.148**	-2.330
ICS 53.554					-3.933	-1.461	-4.512	-6.521	-10.818	-0.411	-0.275	-11.093*
IDS 57.487						2.472	-0.579	-2.588	-6.884	3.521	3.657	-7.159
ICM 55.015							-3.052	-5.060	-9.357	1.050	1.185	-9.632
IDM 58.067								-2.008	-6.304	4.101	4.237	-6.579
CCS 60.075									-4.297	6.109	6.245	-4.572
CDS 64.372										10.407	10.543	-0.275
CCM 53.965											13.612**	-10.682
CDM 53.829												-10.818
Mc 64.641												

[a] mean score for scoring procedure
* 95.0% confidence interval = ± 10.910
**99.0% confidence interval = ∓ 11.510

Table 6.75

Multiple Comparison of Mean Proficiency Scores on CPMP 3 by:
A. Method of Categorization B. Constant Versus Differential Weights C. Single Versus Multiple Keys

ICS	CDM	CCM	ICM	IDS	Author	IDM	CCS	GCS	CDS	Mc	GDS
53.554	53.829	53.965	55.015	57.487	58.054	58.067	60.075	64.291	64.372	64.641	66.978



~~Table~~ 6.76

Multiple Comparison of Mean Errors of Commission Scores on CPMP 3 Calculated Using the 12 Scoring Procedures

	Author	GCS	GDS	ICS	IDS	ICM	IDM	CCS	CDS	CCM	CDM	Mc
	0.541[a]	27.478	25.530	41.141	37.979	42.102	37.472	35.610	24.570	33.249	31.903	28.413
Author 0.541		-26.938**	-24.989**	-40.600**	-37.438**	-41.561**	-36.931**	-35.069**	-24.029**	-32.708**	-31.362**	-28.872**
GCS 27.478			1.948	-13.663**	10.501**	-14.624**	-9.994**	-8.131	2.908	-5.770	-4.425	-1.934
GDS 25.530				-15.611**	-12.449**	-16.572**	-11.942**	-10.080**	0.960	-7.718	-6.372	-3.882
ICS 41.141					3.162	-0.961	3.669	5.531	16.571**	7.892	9.237**	11.728**
IDS 37.979						-4.124	0.507	2.368	13.409**	4.730	6.075	8.566*
ICM 42.102							4.630	6.492	17.532**	8.853*	10.199**	12.689**
IDM 37.472								1.863	12.902**	4.223	5.569	8.059
CCS 35.610									11.039**	2.361	3.706	6.196
CDS 24.570										-8.679*	-7.333	-4.843
CCM 33.249											1.346	-3.835
CDM 31.903												2.490
Mc 29.413												

[a] mean score for scoring procedure

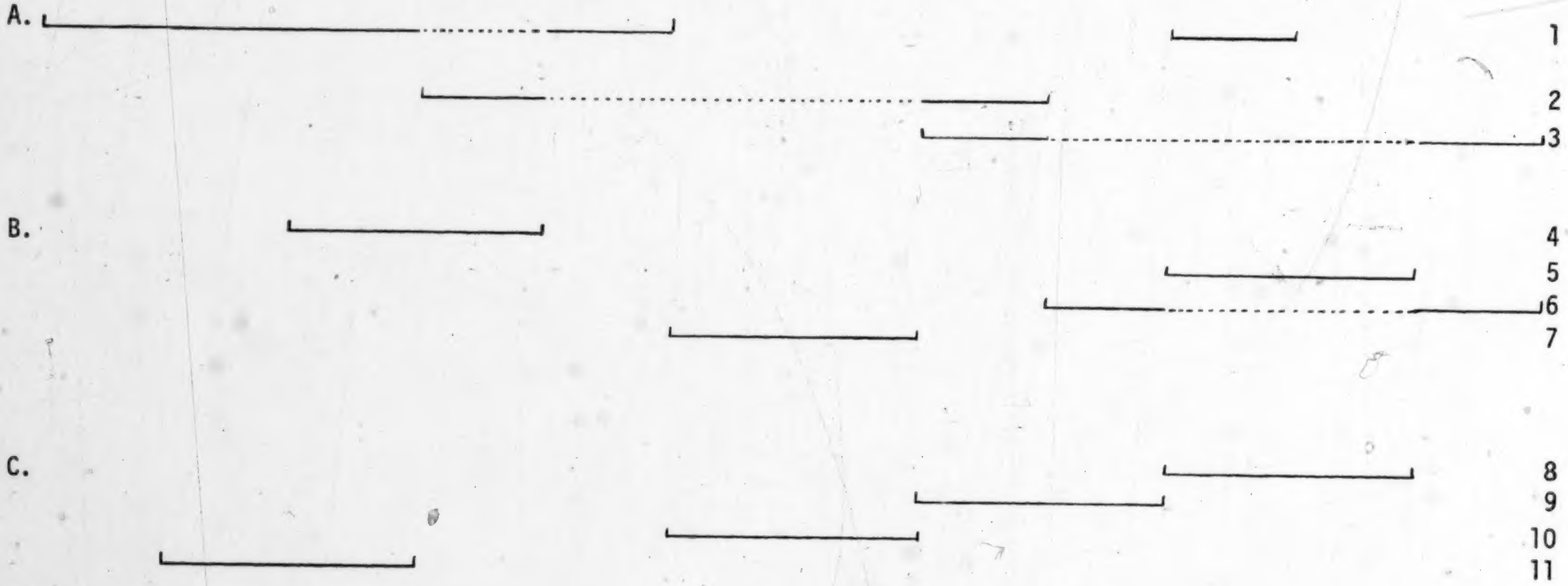
* 95.0% confidence interval = ± 8.417

**99.0% confidence interval = ± 8.879

Table 6.77

Multiple Comparison of Mean Error of Omission Scores on CPMP 3 by:
A. Method of Categorization B. Constant Versus Differential Weights C. Single Versus Multiple Keys

Author	CDS	GDS	GCS	Mc	CDM	CCM	CCS	IDM	IDS	ICS	ICM
24.541	24.570	25.530	27.478	28.413	31.903	33.249	35.610	37.472	37.979	41.141	42.102



of the author, computer, group and McLaughlin methods were found to be equal (see line 1 of Table 6.77).

set 2:

The mean scores of the group and individual methods were statistically different to each other, but both were equal to the mean scores of the computer method (see lines 2-3 of Table 6.77).

ii) constant versus differential weights

No statistical differences were observed among scoring procedures with constant and differential weights (see lines 4-7 of Table 6.77).

iii) single versus multiple keys

No statistical differences were observed among scoring procedures with single and multiple keys (see lines 8-11 of Table 6.77).

d.9) CPMP 3, Efficiency Scores

Table 6.78 presents the multiple comparisons of mean efficiency scores on CPMP 3. Table 6.79 summarizes these results by:

i) method of categorization

set 1:

The mean score of the individual method was significantly larger than the others. There was no significant difference among the mean scores of the author, group and McLaughlin methods. The mean score of the computer method was significantly smaller than the others with the exception of

Table 6.78

Multiple Comparison of Mean Efficiency Scores on CPMP 3 Calculated Using the 12 Scoring Procedures

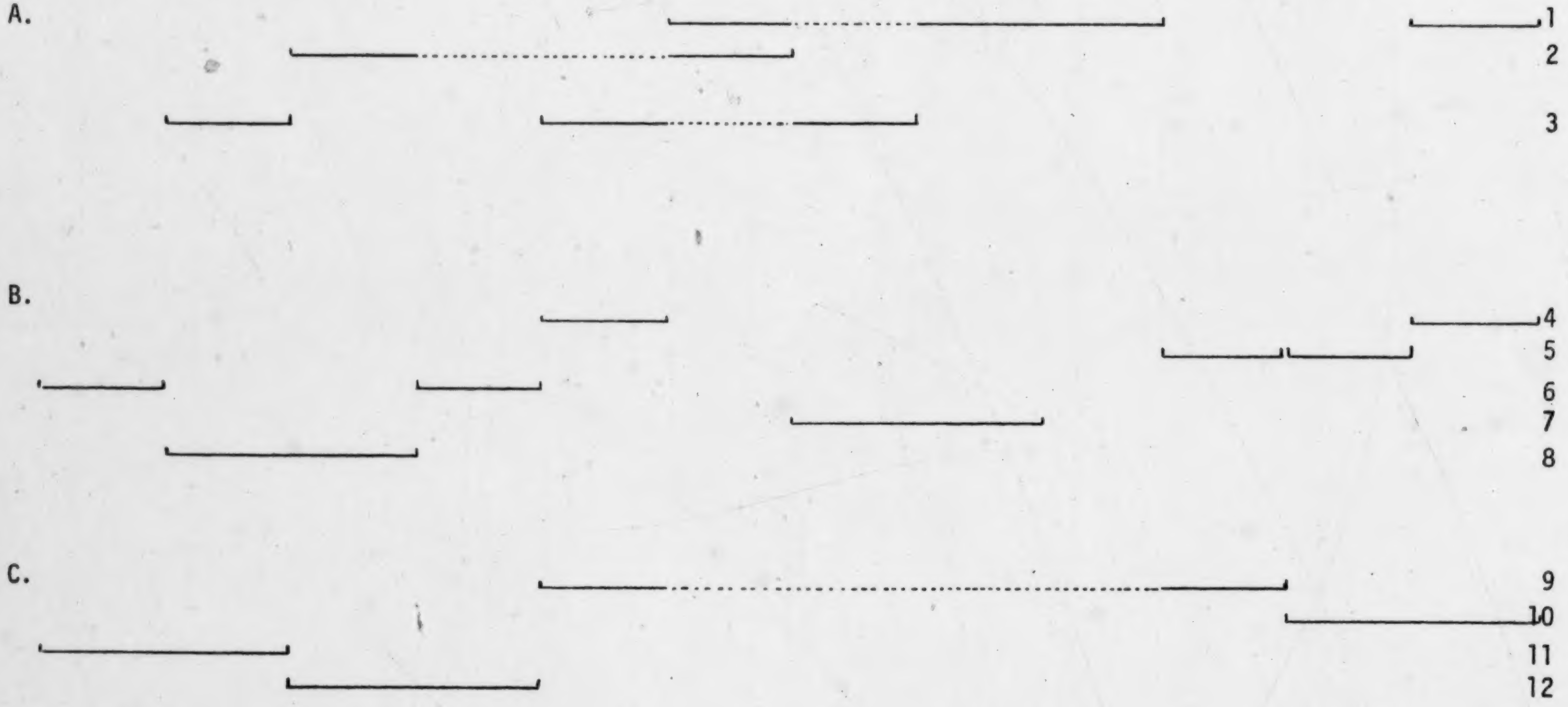
	67.353^{a}	69.115	69.115	66.120	82.032	67.386	80.277	52.221	59.099	49.746	61.893	69.115
Author 67.352		-1.762	-1.762	1.232	-14.680**	-0.034	-12.925**	15.131**	8.252	17.607**	5.458	-1.762
GCS 69.115			0.000	2.994	-12.917**	1.728	-11.162**	16.894**	10.016*	19.370**	7.221	0.000
GDS 69.115				2.994	-12.917**	1.728	-11.167**	16.894**	10.016**	19.370**	7.221	0.000
ICS 66.120					-15.912**	-1.266	-14.157**	13.899**	7.020	16.375**	4.226	-2.994
IDS 82.032						14.646**	1.755	29.811**	22.933**	32.287**	20.139**	12.917**
ICM 67.386							-12.891**	15.165**	8.286	17.641**	5.492	-1.728
IDM 80.277								28.056**	21.177**	30.532**	18.384**	11.162**
CCS 52.221									-6.878	2.475	-9.672	-16.894**
CDS 59.099										9.354	-2.793	10.016*
CCM 49.745											-12.148**	-19.370
CDM 61.893												-7.223
Mc 69.115												

^{a} mean score for scoring procedure
* 95.0% confidence interval = ± 9.774
**99.0% confidence interval = ±10.312

Table 6.79

Multiple Comparison of Mean Efficiency Scores on CPMP 3 by:
A. Method of Categorization B. Constant Versus Differential Weights C. Single Versus Multiple Keys

CCM	CCS	CDS	CDM	ICS	Author	GCS	GDS	Mc	ICM	IDM	IDS
49.746	52.221	59.099	61.893	66.120	67.353	69.115	69.115	69.115	67.386	80.277	82.032



the author mean score (see lines 1-2 of Table 6.79).

set 2:

The mean score of the computer method was significantly lower than the means of the individual and group methods (see line 3 of Table 6.79).

ii) constant versus differential weights

A significant difference in mean scores was observed between the following scoring procedures with constant and differential weights: IDS > ICS, IDM > ICM and CDM > CCM; but the mean scores of the following comparisons were found equal: GCS = GDS and CCS = CDS.

iii) single versus differential weights

No statistical differences were observed between mean efficiency scores.

d.10 CPMP 4, Proficiency Scores

Table 6.80 presents the multiple comparisons of mean proficiency scores on CPMP 4. Table 6.81 summarizes these results by:

i) method of categorization

set 1:

The mean score of the author scoring procedure was significantly lower than the others. There was no significant difference among the mean scores of the McLaughlin, individual, group and computer methods (see line 1 of Table 6.81).

set 2:

The mean score of the group method

Table 6.80

Multiple Comparison of Mean Proficiency Scores on CPMP 4
Calculated Using the 12 Scoring Procedures

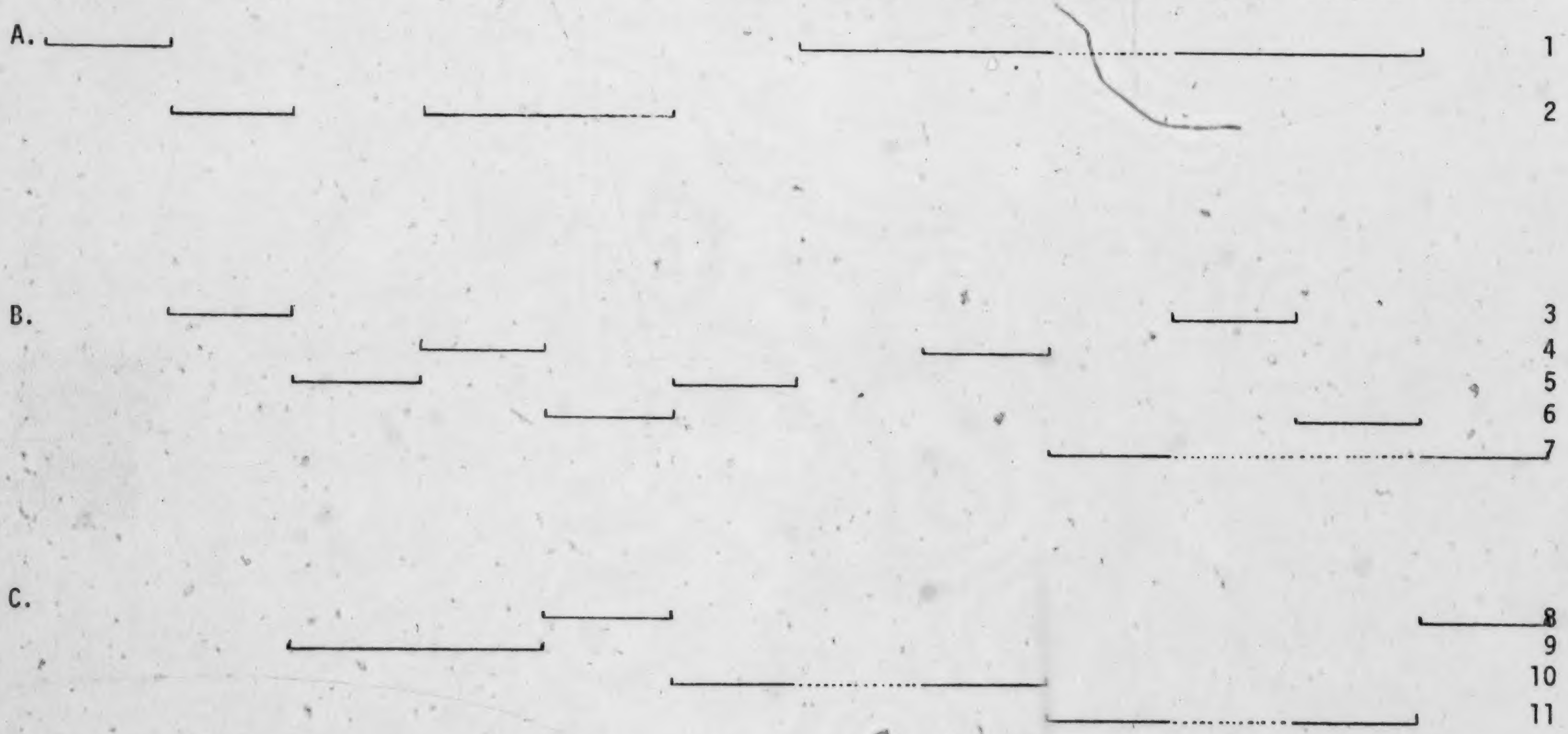
	Author	GCS	GDS	ICS	IDS	ICM	IDM	CCS	CDS	CCM	CDM	Mc
	75.400^a	76.347	85.248	79.194	83.434	77.379	81.428	79.668	85.647	86.036	84.242	82.396
Author 75.400		-0.947	-9.848**	-3.794	-8.034**	-1.979	-6.028**	-4.268**	-10.247**	-10.636**	-8.842**	-6.996**
GCS 76.347			-8.901**	-2.847	-7.087**	-1,032	-5.081**	-3.321	-9.300**	-9.689**	-7.895**	-6.049**
GDS 85.248				6.054**	1.814	7.869**	3.820	5.580**	-0.399	-0.788	1.006	2.852
ICS 79.194					-4.240**	1.815	-2.234	-0.474	-6.453**	-6.842**	-5.048**	-3.202
IDS 83.434						6.054**	2.006	3.766	2.213	-2.602	-0.808	1.038
ICM 77.379							-4.049*	-2.289	-8.268**	-8.657**	-6.863**	-5.017**
IDM 81.428								1.759	-4.219**	-4.608**	-2.814	-0.968
CCS 79.668									-5.979**	-6.368**	-4.574**	-2.728
CDS 85.647										-3.887	1.405	3.251
CCM 86.036											1.794	3.640
CDM 84.242												1.846
Mc 82.396												

[a] mean score for scoring procedure
* 95.0% confidence interval = ± 3.907
**99.0% confidence interval = ± 4.121

Table 6.81

Multiple Comparison of Mean Proficiency Scores on CPMP 4 by:
A. Method of Categorization B. Constant Versus Differential Weights C. Single Versus Multiple Keys

Author	GCS	ICM	ICS	CCS	IDM	Mc	IDS	CDM	GDS	CDS	CCM
75.400	76.347	77.379	79.194	79.668	81.428	82.396	83.434	84.242	85.248	85.647	86.036



was significantly different from the individual and computer mean scores (see line 2 of Table 6.81).

ii) constant versus differential weights

A statistical difference was observed between four of the five comparisons: GDS > GCS, IDS > ICS, IDM > ICM and CDS > CCS; but no difference was found between CCM and CDM (see lines 3-7 of Table 6.81).

iii) single versus multiple keys

Only one comparison was statistically significant: CCM > CCS; but no difference was found between the other three paired comparisons: ICS = ICM, IDS = IDM and CDS = CDM.

d.11) CPMP 4, Error of Omission Scores

Table 6.82 presents the multiple comparisons of mean error of omission scores on CPMP 4. Table 6.83 summarizes these results by:

i) method of categorization

set 1:

The mean score of the computer method was significantly lower than the other means which were observed to be equal (see line 1 of Table 6.83).

set 2:

No statistical differences were observed among the mean scores of the individual, group and computer methods (see line 2 of Table 6.83)

ii) constant versus differential weights

Table 6.82

Multiple Comparison of Mean Errors of Commission Scores on CPMP 4 Calculated Using the 12 Scoring Procedures

	15.582[a]	19.590	12.504	18.574	14.197	20.401	16.095	19.409	6.793	11.411	7.590	13.949
Author 15.582		-3.988**	3.078**	-2.992*	1.385	-4.819**	-0.509	-3.827**	8.789**	4.171**	7/992**	1.633
GCS 19.570			7.066**	0.996	5.373**	-0.831	3.479**	0.161	12.777**	8.159**	11.980**	5.621**
GDS 12.504				-6.070**	-1.693	-7.897**	-3.587**	-6.905**	5.711**	1.093	4.914**	-1.445
ICS 18.574					4.377**	-1.827	2.483	-0.835	11.781**	7.163**	10.984**	4.625**
IDS 14.197						-6.204**	-1.894	-5.212**	7.404**	2.786	6.607	0.248
ICM 20.401							4.311**	0.992	13.608**	8.990**	12.811**	6.452**
IDM 16.091								-3.319**	9.298**	4.680**	8.501**	2.142
CCS 19.409									12.616**	7.998**	11.819**	5.460**
CDS 6.793										-4.618**	-0.797	7.156**
CCM 11.411											3.821**	2.538
CDM 7.590												6.359**
Mc 13.949												

[a] mean score for scoring procedure

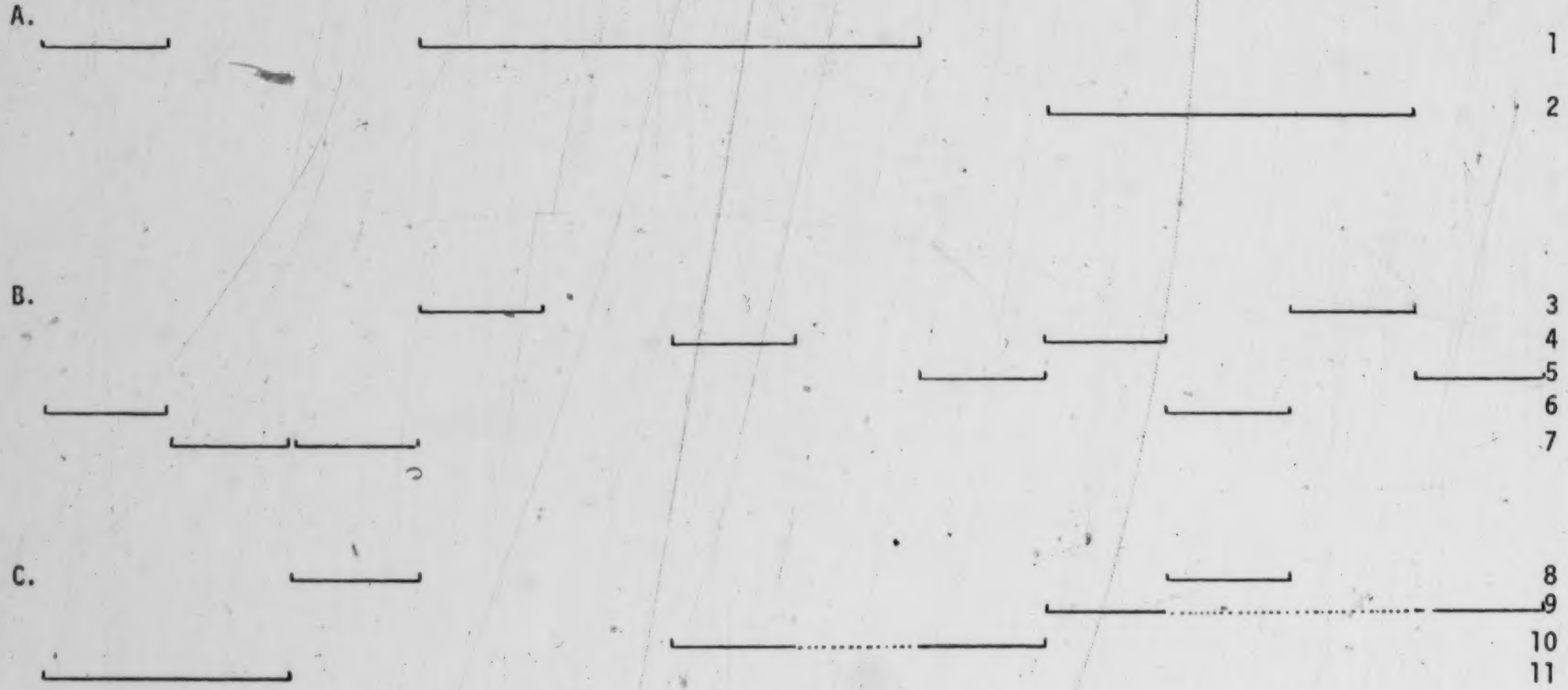
* 95.0% confidence interval = ± 2.838

**99.0% confidence interval = ± 2.994

Table 6.83

Multiple Comparison of Mean Error of Omission Scores on CPMP 4 by:
A. Method of Categorization B. Constant Versus Differential Weights C. Single Versus Multiple Keys

CDS	CDM	CCM	GDS	Mc	IDS	Author	IDM	ICS	CCS	GCS	ICM
6.793	7.590	11.411	12.504	13.949	14.197	15.582	16.095	18.574	18.409	19.590	20.401



Mean scores of scoring procedures with differential weights were statistically lower than those with constant weights: GCS > GDS, ICS > IDS, ICM > IDM, CCS > CDS and CCM > CDM (see lines 3-7 of Table 6.83).

iii) single versus multiple keys

Only one comparison was observed to be significantly different: CCS > CCM. No differences in mean scores were observed among the remaining scoring procedures: ICS = ICM, IDS = IDM and CDS = CDM (see lines 8-11 of Table 6.83).

d.12) CPMP 4, Efficiency Scores

Table 6.84 presents the multiple comparisons of mean efficiency scores on CPMP 4. Table 6.85 summarizes these results by:

i) method of categorization

set 1:

The mean scores of the author and computer methods were statistically lower than those of the McLaughlin and group methods which in turn were lower than the individual method (see line 1 of Table 6.85).

set 2:

No statistical difference was found among the mean scores calculated by scoring procedures with constant weights and single keys: ICS = GCS = CCS (see line 2 of Table 6.85).

ii) constant versus differential weights

The following statistical differences in mean efficiency scores with constant versus differential

Table 6.84

Multiple Comparison of Mean Efficiency Scores on CPMP 4 Calculated Using the 12 Scoring Procedures

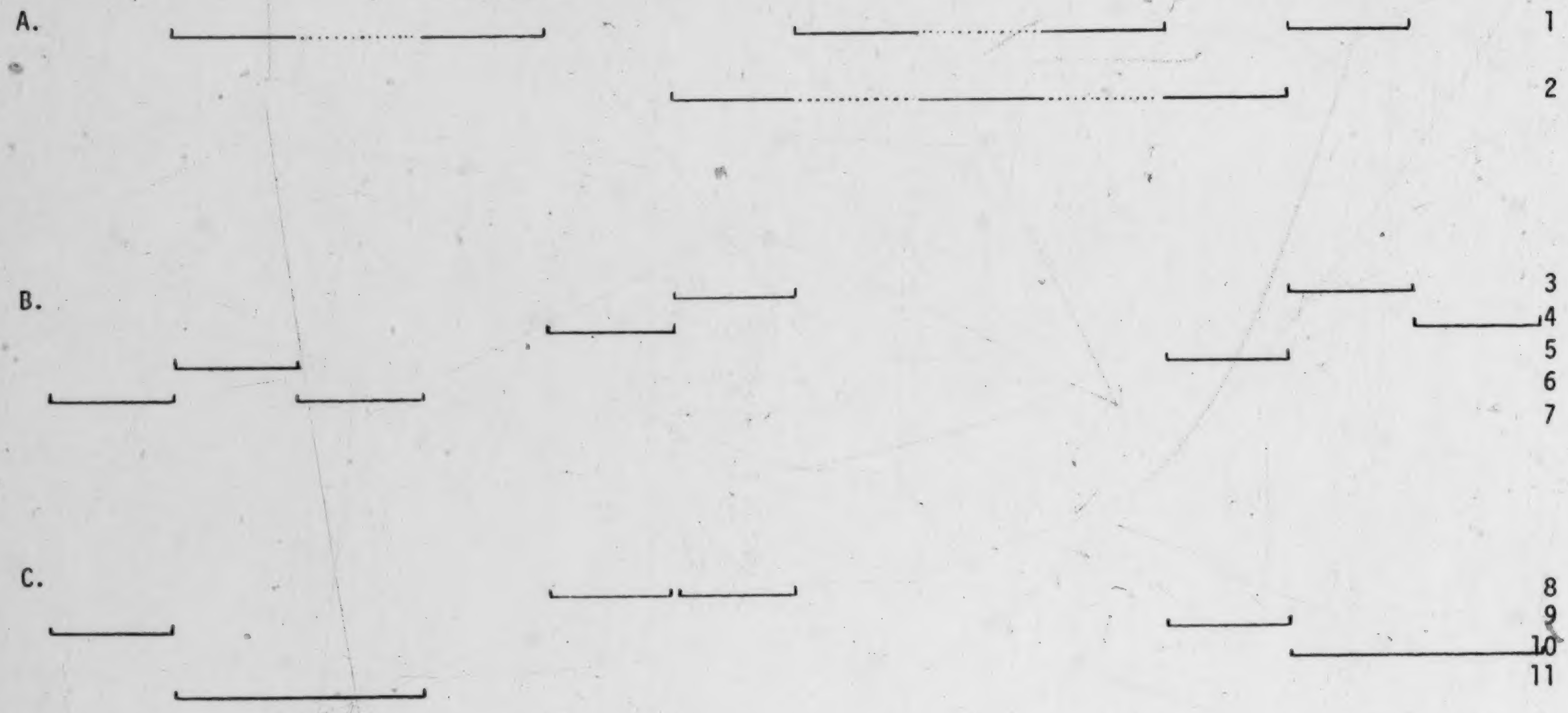
	Author	GCS	GDS	ICS	IDS	ICM	IDM	CCS	CDS	CCM	CDM	Mc
	1.824[a]	83.302	83.302	81.336	89.441	74.872	90.373	85.307	71.123	65.908	71.242	83.036
Author 71.824		-11.478**	-11.478**	-9.512**	-17.487**	-3.048	-18.549**	-13.483**	0.701	5.916**	0.582	-11.212**
GCS 83.302			0.000	1.966	-6.109**	8.430**	07.071**	-2.005	12.179**	17.394**	12.060**	0.266
GDS 83.302				1.965	-6.109**	8.430**	-7.071**	-2.005	12.179**	17.394**	12.060**	0.266
ICS 81.336					-8.075**	-8.075**	6.464**	-9.037**	-3.971	10.213**	10.094**	-1.700
IDS 89.411						14.540**	-0.962	4.104	18.288**	23.503**	18.169**	6.375**
ICM 74.872							-15.501**	-10.435**	3.749	8.964	3.630	-8.164**
IDM 90.373								5.066**	19.250**	24.465**	19.131**	7.337**
CCS 85.307									14.184**	19.399**	14.065**	2.271
CDS 71.123										5.215**	0.119	-11.913**
CCM 65.908											-5.334**	-17.128**
CDM 71.242												-11.794**
Mc 83.036												

[a] mean score for scoring procedure
* 95.0% confidence interval = ± 4.625
**99.0% confidence interval = ± 4.880

Table 6.85

Multiple Comparison of Mean Efficiency Scores on CPMP 4 by:
A. Method of Categorization B. Constant Versus Differential Weights C. Single Versus Multiple Keys

CCM	CDS	CDM	Author	ICM	ICS	Mc	GCS	GDS	CCS	IDS	IDM
65.908	71.123	71.242	71.824	74.872	81.336	83.036	83.302	83.302	85.307	89.441	90.373



weights: IDS > ICS, IDM > ICM, CCS > CDS and CDM > CCM (see lines 3-7 of Table 6.85).

iii) single versus multiple keys

The following statistical differences in mean efficiency scores were observed between scoring procedures of single and multiple keys: ICS > ICM, CCS > CCM; but no differences were observed between the means of the following scoring procedures: IDS = IDM and CDS = CDM (see lines 8-11 of Table 6.85).

E. Summary and Discussion of Results Obtained in the One-Way Multivariate Analysis

The summary and discussion of the effects of scoring procedures upon examinee CPMP mean scores has been divided into three sub-topics, namely, effects due to i) method of categorization, ii) constant versus differential weights, and iii) single versus multiple keys.

a) Effects Due to Method of Categorization

Tables 6.86, 6.87, 6.88, 6.89, 6.90 and 6.91 respectively summarize the multiple comparisons of mean proficiency, error of omission and efficiency scores. Statistically equal means are identified by a common underline. An examination of the tables yields the following observations:

1) there were no consistent changes in mean scores over the four CPMPs

Table 6.86

Summary of Multiple Comparison of Mean Proficiency Scores on CPMP 1-4 Calculated Using Scoring Procedures With Differential Weights and Single Keys:

CPMP					
1	AUTHOR	GDS	Mc	IDS	CDS
	85.357	85.673	86.401	87.488	89.131
2	AUTHOR	IDS	GDS	Mc	CDS
	81.434	82.700	84.512	85.084	89.528
3	IDS	AUTHOR	CDS	Mc	GDS
	57.487	58.054	64.372	64.647	66.978
4	AUTHOR	Mc	IDS	GDS	CDS
	75.400	82.396	83.434	85.248	85.647

Table 6.87

Summary of Multiple Comparison of Mean Proficiency Scores on CPMP 1-4 Calculated Using Scoring Procedures with Constant Weights and Single Keys:

CPMP			
1	GCS	ICS	CCS
	81.931	83.234	84.347
2	ICS	GCS	CCS
	80.568	80.814	86.439
3	ICS	CCS	GCS
	53.554	60.075	64.291
4	GCS	ICS	CCS
	76.347	79.194	79.668

Table 6.88

Summary of Multiple Comparison of Mean Error of Omission Scores on CPMP 1-4 Calculated Using Scoring Procedures with Differential Weights and Single Keys:

CPMP					
1	CDS 5.239	AUTHOR 10.317	IDS 10.556	Mc 10.591	GDS 11.528
2	CDS 5.108	Mc 8.945	AUTHOR 10.047	GDS 10.082	IDS 14.863
3	AUTHOR 0.540	CDS 24.570	GDS 25.530	Mc 29.413	IDS 37.979
4	CDS 6.793	GDS 12.504	Mc 13.949	IDS 14.197	AUTHOR 15.582

Table 6.89

Summary of Multiple Comparison of Mean Error of Omission Scores on CPMP 1-4 Calculated Using Scoring Procedures with Constant Weights and Single Keys:

CPMP			
1	ICS 14.110	CCS 14.339	GCS 14.578
2	GCS 9.479	CCS 12.723	ICS 16.058
3	GCS 27.478	CCS 35.610	ICS 41.141
4	ICS 18.574	CCS 19.409	GCS 19.570

Table 6.90

Summary of Multiple Comparison of Mean Efficiency Scores on CPMP 1-4 Calculated Using Scoring Procedures with Differential Weights and Single Keys:

CPMP					
1	CDS 81.481	AUTHOR 84.799	Mc 87.267	IDS 87.390	GDS 87.969
2	Mc 72.884	GDS 73.154	CDS 73.671	AUTHOR 78.477	IDS 88.931
3	CDS 59.099	AUTHOR 67.352	GDS 69.115	Mc 69.115	IDS 82.032
4	AUTHOR 71.824	CDS 71.123	Mc 83.036	GDS 83.302	IDS 89.411

Table 6.91

Summary of Multiple Comparison of Mean Efficiency Scores on CPMP 1-4 Calculated Using Scoring Procedures with Constant Weights and Single Keys:

CPMP			
1	CCS 83.636	IDS 87.896	GCS 87.969
2	GCS 73.154	CCS 77.943	ICS 81.424
3	CCS 52.221	ICS 66.120	GCS 69.115
4	ICS 81.336	GCS 32.302	CCS 85.307

2) there were significant differences in mean scores within CPMPs 1, 2 and 4 but few in CPMP 3

3) the author's scoring procedure tended to produce the lowest mean proficiency scores while the computer scoring procedure tended to produce the highest

4) the computer scoring key with differential weights tended to yield the lowest mean error of omission scores

5) the individual scoring key tended to yield the highest mean efficiency score

6) the mean proficiency, error of omission and efficiency scores of the McLaughlin and GDS scoring procedures were statistically equal on all CPMPs.

Each of the above six observations will be discussed in turn.

Firstly, no consistent pattern was evident throughout the above tables. The lack of a consistent pattern may be due to differences among groups of experts or to differences in CPMPs, or to both. The groups of experts may have varied in terms of their medical knowledge and experiences although it has been earlier assumed that the composition of the groups was equal. The CPMPs varied in terms of the nature of the presenting medical problem; the type of care required; the emphasis on history, physical examination, laboratory investigation and management; and, the type of branching employed. It is believed that the lack of a consistent pattern was primarily due to the differences in CPMPs, however, differences in the groups of experts may have had a minor effect.

Secondly, the variation in the number of significant paired comparisons among the CPMP means is related to the size of the confidence interval. Table 6.92 presents both the number of significant paired comparisons out of the total of 66, and the

Table 6.92

Number of Significant Paired Comparisons of Mean Proficiency, Error of Omission, and Efficiency Scores on CPMPs 1-4 and the Corresponding 95% Confidence Interval

	CPMP			
	1	2	3	4
Proficiency	40(±1.954)*	39(±2.765)	6(±10.910)	33(±3.907)
Error of Omission	43(±1.368)	51(±1.768)	26(± 8.417)	46(±2.838)
Efficiency	45(±2.631)	46(±3.377)	35(± 9.774)	47(±4.625)

*95% confidence interval

size of the 95% confidence interval. CPMP 3, the most complex branching problem used in this study, had the fewest number of significant differences and the largest confidence interval. CPMP 3 also had the largest amount of variation of opinions, judgements and selections when the group, individual and computer methods were respectively employed to categorize options. This variation occurred both within and across methods and was reflected in changes to option weights and optimal pathways. The consequence of this variation was an increased error term which was three to five times that of other scoring procedures and few significant differences among mean scores.

Thirdly, it was observed that the author's mean proficiency score tended to be the lowest while the computer's was the highest (see Table 6.86). These results may be explained by comparing the scoring keys to the frequency with which options were selected. This was accomplished in two ways:

1) by comparing the number of positive options identified within the eight scoring keys (see Table 6.93) to the

average number of selections made by experts and examinees, (see Table 6.93), and

2) by calculating correlation coefficients between scoring key option weightings and the frequency with which options were selected by experts and examinees (see Table 6.94).

In the first comparison, it was observed that both examinees and expert problem-solvers tended to select fewer options than those categorized as positive by author, group and individual keys (see Table 6.93). Since the author and the other scoring procedures identified more positive options than those selected by examinees, this lowered the examinee mean proficiency scores and increased their error of omission scores. However, this phenomenon did not occur with the computer scoring procedure since there was a closer correspondence between options categorized as positive and the number of positive options selected by the examinees.

In the second comparison it was observed that the correlation coefficients for the author method were the lowest while those for the computer method tended to be the highest (see Table 6.94). Thus the computer scoring key produced the highest mean proficiency scores because examinee selections closely matched the weightings of the computer scoring key.

Fourthly, the computer scoring key with differential weights and a single key (CDS) tended to yield the lowest mean error of omission score (see Table 6.89). As noted earlier, mean scores were directly related to option categorization and frequency of selection. The closer the above match, the higher

Table 6.93

Number of Positive Options Identified Within Eight Scoring Procedures and Average Number of Selections by Examinee and Expert Problem-Solver

CPMP	Scoring Procedures									
	Author	GDS	IDS	CDS	Mc	GCS	ICS	CCS	Expert	Examinee
1	39*	43	44	36	43	43	41	37	36**	34
2	36	33	46	24	33	33	37	26	24	22
3	20	13	17	15	13	13	14	12	15	15
4	29	30	28	23	30	30	28	24	23	21

* number of positive options identified in scoring procedures

** average number of selections

Table 6.94

Correlation Coefficient Between Scoring Key and Frequency that Options were Selected in CPMPs 1-4 by Expert Problem-Solvers and Examinees

	CPMP 1		CPMP 2		CPMP 3		CPMP 4	
	Expert	Examinee	Expert	Examinee	Expert	Examinee	Expert	Examinee
Author	0.78	0.84	0.56	0.61	0.23	0.37	0.50	0.45
GDS	0.80	0.83	0.79	0.74	0.70	0.66	0.88	0.83
IDS	0.90	0.91	0.60	0.64	0.67	0.63	0.80	0.76
CDS	1.00	0.92	1.00	0.92	0.85	0.62	0.96	0.86
Mc	0.91	0.92	0.78	0.76	0.81	0.79	0.86	0.77
GCS	0.77	0.84	0.59	0.64	0.65	0.69	0.63	0.55
ICS	0.83	0.84	0.54	0.58	0.62	0.56	0.72	0.66
CCS	0.90	0.83	0.92	0.85	0.90	0.66	0.85	0.80
Expert	1.00	0.92	1.00	0.92	1.00	0.74	1.00	0.91

the mean proficiency score and the lower the mean error of omission score. If the average number of selections was lower than the number of options categorized as positive, the mean error of omission increased. Therefore, the relatively low mean error of omission scores produced by the computer method were due to the close match between options categorized as positive and selections by examinees.

Fifthly, it was observed that the individual method tended to yield the highest mean efficiency score (see Table 6.90). This observation can be explained by examining the formula used to calcualte the efficiency score. Efficiency is defined as the percentage of positive options selected over the total number of options selected (i.e., Efficiency % = number of positive selections X 100/total number of selections). Since the individual method tended to have the largest number of options categorized as positive, the number of positive selections made by the examinee increased compared to the total number of selections, thereby producing the relatively higher mean efficiency scores for the individual method.

Lastly, it was observed that the mean proficiency, error of omission and efficiency scores of the McLaughlin and GDS scoring procedures were statistically equal on all CPMPs. The only difference between the two procedures was the method used to assign differential weights. In the group procedure, the weights reflected the collective judgements of the group of experts, while in the McLaughlin procedure, they reflected the collective expert problem-solvers' selections. The two methods of differential weightings however had no effect upon altering

CPMP mean scores.

b) Effect of Constant and Differential Weights Upon Examinee CPMP Mean Scores

Table 6.95 summarizes the results of constant and differential weights upon CPMP mean scores. The table provides the results by scoring procedure (i.e., GCS vs GDS) and overall (i.e., GCS + ICS + ICM + CCS + CCM vs GDS + IDS + IDM + CDS + CDM). Although the results are not consistent, they do seem to indicate that:

1) mean proficiency scores were larger for scoring procedures with differential weights

2) mean error of omission scores were larger for scoring procedures with constant weights, and

3) mean efficiency scores were larger for the individual scoring procedure employing differential weights although, overall, the effect was insignificant.

These results will be discussed in turn.

Firstly, the observation of higher proficiency scores for scoring procedures with differential weights may be explained by examining the following formula:

$$\bar{P}_j = \frac{1}{N} \sum_{i=1}^{I} f_i w_{ij} \qquad (6.3)$$

where $\bar{P}_j$ = the mean true/false proficiency score calculated using the jth scoring procedure

N = the number of examinees

f_i = the frequency of either selection or non-selection (i.e., if option i is positive, then f_i is the frequency of selection, but if option i is negative, then f_i is the frequency of non-selection)

Table 6.95

Summary of Multiple Comparison of Mean Scores on CPMP 1-4
Calculated Using Scoring Procedures With Constant and Differential Weights

	CPMP											
	1			2			3			4		
	SCORE											
	Prof	EofO	Eff	Prof	EofO	Eff	Prof	EofO	Eff	Prof	EofO	Eff
GCS vs GDS	<	>	NS	<	NS	NS	NS	NS	NS	<	>	NS
ICS vs IDS	NS	>	NS	NS	NS	<	NS	NS	<	<	>	<
ICM vs IDM	NS	NS	<	NS	NS	<	NS	NS	<	<	>	<
CCS vs CDS	<	>	NS	<	>	>	NS	NS	NS	<	>	>
CCM vs CDM	NS	>	NS	NS	>	>	NS	>	<	NS	>	<
Overall	NS	>	<	<	>	NS	NS	NS	NS	<	>	NS

Prof = Proficiency Score
EofO = Error of Omission
Eff = Efficiency Score
NS = not significant
< = less than (*i.e.*, GCS < GDS)
> = greater than (*i.e.*, ICS > IDS)

w_{ij} = the absolute weighting assigned option i by the scoring procedure

In scoring procedures with constant weights, w_{ij} is equal for all options but in scoring procedures with differential weights, w_{ij} varies according to the option's perceived degree of importance. When constant weights are used, equation 6.3 can be re-written as follows:

$$\bar{P}_j = \frac{w_j}{N} \sum_{i=1}^{I} f_i \qquad (6.4)$$

With the constant, w_j, moved outside the summation sign, the magnitude of $\bar{P}_j$ is dependent upon the frequency with which positive options are selected and negative options avoided. $\bar{P}_j$ will be maximized when all examinees select positive options and avoid negative ones. When differential weights are used, and f_i and w_{ij} tend to vary together (i.e., when f_i is large, w_{ij} is large), the total and the mean will tend to increase over that of constant weights. Thus, the mean proficiency scores were increased by the differential weights because f_i and w_{ij} tended to vary together.

The explanation for the second observation, that mean error of omission scores were larger for constant weights, is the converse of that explanation given for the increase in mean proficiency scores. If the mean proficiency scores of the differentially weighted scoring procedures were larger, the mean omission scores were smaller. The converse is likewise true. Thus the error of omission scores for constant weights were larger because the mean proficiency scores were smaller.

Lastly, efficiency scores were also affected by the differential weights in the individual scoring procedure which resulted in higher scores. In the individual scoring procedures employing differential weightings, there was an increase in the number of positive options therefore an increase in mean efficiency scores.

c) Effect of Single and Multiple Keys upon Examinee CPMP Mean Scores

Table 6.96 summarizes the effect of single and multiple keys upon examinee CPMP mean scores. The results indicate that there is no consistent overall effect upon examinee scores. This finding reflects the difficulty encountered in identifying more than one homogeneous sub-group among small groups of experts. There were only 10, 16 and 11 participants respectively in groups A, B and C. The techniques employed were unsuccessful in identifying homogeneous sub-groups within these groups. Only the computer procedure in CPMP 3 yielded more than one homogeneous sub-group. However, the error term in CPMP 3 was so large that no differences were found among the multiple comparisons of mean scores.

The effect of scoring procedures upon examinee satisfactory (pass)/unsatisfactory (fail) states is examined in the next section.

5. Number of Examinees Who Achieved Satisfactory Status on CPMP 1-4

Table 6.97 presents the number of examinees who were

Table 6.96

Summary of Multiple Comparison of Mean Scores on CPMP 1-4 Calculated Using Scoring Procedures With Single and Multiple Keys

	CPMP											
	1			2			3			4		
	SCORE											
	Prof	EofO	Eff	Prof	EofO	Eff	Prof	EofO	Eff	Prof	EofO	Eff
ICS vs ICM	NS	>	>	NS	NS	NS	NS	NS	NS	NS	NS	>
IDS vs IDM	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
CCS vs CCM	<	>	>	NS	NS	NS	NS	NS	NS	<	>	>
CDS vs CDM	NS	>	>	NS	NS	NS	NS	NS	NS	NS	NS	>
Overall	NS	>	>	NS	NS	NS	NS	NS	NS	NS	NS	>

Prof = Proficiency Score
EofO = Error of Omission
Eff = Efficiency Score
NS = not significant
< = less than (i.e., CCS CCM)
> = greater than (i.e., ICS ICM)

Table 6.97

Number of Examinees (N = 111) Who Achieved Satisfactory Status
on CPMPs 1-4 Scored by 12 Different Scoring Procedures

	CPMP			
	1	2	3	4
MPL	75%	70%	60%	70%
Author	98	71	70	67
GCS	93	70	58	89
GDS	98	88	87	107
ICS	95	65	35	101
IDS	98	80	39	108
ICM	98	53	52	90
IDM	98	72	44	106
CCS	90	102	53	93
CDS	110	106	81	111
CCM	109	100	13	110
CDM	110	106	12	111
Mc	97	99	80	108

declared satisfactory (*i*.*e*., score $\geq$ MPL) on CPMPs 1-4 scored by twelve different scoring procedures. The numbers varied from 90 to 110 on CPMP 1, from 53 to 106 on CPMP 2, from 12 to 87 on CPMP 3, and, from 67 to 111 on CPMP 4. The largest discrepancies occurred in CPMPs 2, 3 and 4. CPMP 2 was selected for further analysis in order to determine whether the change in examinee status among scoring procedures occurred due to a shift from satisfactory to unsatisfactory or visa versa. This analysis was deemed necessary since shifts could have occurred when there was no or little difference in absolute numbers. For example, there were 71 and 70 students respectively declared satisfactory in CPMP 2 using the author and GCS scoring procedures. Were the same examinees declared satisfactory and unsatisfactory or did the status of only certain examinees change? Table 6.98 presents the disagreements in status among examinees by scoring procedures. The diagonal numbers represent the number of examinees whose performance was either satisfactory (pass)/unsatisfactory (fail) on CPMP 2 by scoring procedure. The off-diagonal numbers represent the number of shifts. For example, nine examinees that were declared satisfactory by the author's scoring procedure were unsatisfactory by the GCS procedure. Likewise, 8 examinees declared unsatisfactory by the author's scoring procedure were declared satisfactory by the GCS procedure. An examination of Table 6.98 would reveal that as few as zero (_i.e._, see scoring procedure CDS and CDM) and as many as 53 (_i.e._, see scoring procedures ICM and CDS) disagreements occurred.

On the basis of the above observations, it was concluded that examinee satisfactory/unsatisfactory status could be altered depending upon the CPMP, scoring procedure and minimum pass level (MPL).

The degree to which different scoring procedures altered examinees' rank ordering is reported in the next section.

6. Rank Ordering of Examinee Scores

Tables 6.99 to 6.102 present rank ordering (Spearman's rho) of examinee scores calculated respectively on CPMPs 1-4 using the twelve scoring procedures There was a large variation in coefficients among scoring procedures. From the tables it can be observed that the coefficients varied from 0.56 to 0.98 in CPMP 1, 0.33 to 0.99 in CPMP 2, -0.11 to 0.90 in CPMP 3, and 0.19 to 1.00 in CPMP 4. Using the statistical test of $\rho_{xy} - \rho_{xz}$ for dependent

Table 6.98
Number of Disagreements in Satisfactory (Pass)/ Unsatisfactory (Fail) Status on CPMP 2 by Scoring Procedure

AUTHOR	GCS	GDS	ICS	IDS	ICM	IDM	CCS	CDS	CCM	CDM	Mc
71/40	9*/8**	3/20	17/11	9/18	25/7	14/15	3/34	0/35	4/33	0/35	0/28
	70/41	0/18	18/13	8/18	26/9	17/19	3/35	0/36	4/34	0/36	0/29
		88/23	27/4	5/13	36/1	23/7	6/20	3/21	7/19	3/21	1/12
			65/36	1/16	12/0	4/11	2/39	1/42	3/38	1/42	1/35
				80/37	27/6	10/2	3/25	2/28	4/24	2/28	1/20
					53/58	1/20	1/50	0/53	1/48	0/53	1/47
						72/39	4/34	2/36	5/33	2/36	2/29
							102/9	0/4	2/0	0/4	8/5
								106/5	6/0	0/0	9/2
									100/11	0/6	8/7
										106/5	9/2
											99/12

* number of satisfactory changed to unsatisfactory
** number of unsatisfactory changed to satisfactory

Table 6.99

Rank Ordering (Spearman's Rho*) of Students (N = 111) By Proficiency Scores Calculated on CPMP 1 Using 12 Scoring Procedures

	Author	GCS	GDS	ICS	IDS	ICM	IDM	CCS	CDS	CCM	CDM	Mc
Author	100	76	81	72	74	79	79	56	69	61	73	78
GCS		100	92	84	82	82	83	77	57	76	63	96
GDS			100	89	90	89	92	79	69	78	73	94
ICS				100	94	90	94	85	71	85	73	89
IDS					100	94	98	88	71	87	72	90
ICM						100	96	83	68	83	72	87
IDM							100	86	74	87	76	91
CCS								100	61	94	63	85
CDS									100	67	97	71
CCM										100	74	86
CDM											100	76
Mc												100

*decimal point omitted

Table 6.100

Rank Ordering (Spearman's Rho*) of Students (N = 111) by Proficiency Scores

Calculated on CPMP 2 Using 12 Scoring Procedures

	Author	GCS	GDS	ICS	IDS	ICM	IDM	CCS	CDS	CCM	CDM	Mc
Author	100	86	84	65	59	63	57	36	57	44	64	83
GCS		100	91	60	51	58	47	34	62	33	65	93
GDS			100	71	66	70	64	40	55	38	58	89
ICS				100	96	99	95	67	50	67	52	69
IDS					100	96	99	69	61	70	51	64
ICM						100	96	67	48	64	49	67
IDM							100	68	47	69	48	60
CCS								100	67	93	63	60
CDS									100	72	97	79
CCM										100	71	60
CDM											100	80
Mc												100

*decimal point omitted

Table 6.101

Rank Ordering (Spearman's Rho*) of Students (N = 111) By Proficiency Scores

Calculated on CPMP 3 Using 12 Scoring Procedures

	Author	GCS	GDS	ICS	IDS	ICM	IDM	CCS	CDS	CCM	CDM	Mc
Author	100	34	53	46	49	60	48	42	20	-11	-05	30
GCS		100	70	56	56	55	58	40	40	28	26	78
GDS			100	56	54	62	57	42	36	21	25	58
ICS				100	84	83	85	51	31	08	03	50
IDS					100	82	90	59	36	07	05	52
ICM						100	86	62	33	11	10	50
IDM							100	56	33	11	10	52
CCS								100	58	25	24	53
CDS									100	35	41	61
CCM										100	77	29
CDM											100	29
Mc												100

*decimal point omitted

Table 6.102

Rank Ordering (Spearman's Rho*) of Students (N = 111) By Proficiency Scores Calculated on CPMP 4 Using 12 Scoring Procedures

	Author	GCS	GDS	ICS	IDS	ICM	IDM	CCS	CDS	CCM	CDM	Mc
Author	100	90	83	73	73	64	74	69	50	38	33	81
GCS		100	91	83	79	70	79	73	43	42	27	88
GDS			100	83	73	63	73	73	34	35	19	77
ICS				100	91	85	90	88	55	62	45	87
IDS					100	96	100	85	75	81	70	92
ICM						100	96	76	81	86	78	89
IDM							100	84	75	80	70	92
CCS								100	56	67	47	80
CDS									100	70	81	55
CCM										100	75	59
CDM											100	45
Mc												100

*decimal point omitted

samples, it was necessary for two correlation coefficients to differ by approximately 0.10 before they were considered significantly different at the 0.01 level. Applying this general guideline to the correlation coefficients in Tables 6.99 to 6.102, it is evident that the scoring procedures did significantly alter many of the rank orderings of examinees in all four CPMPs. From this data it was concluded that different scoring procedures did alter examinee rank orderings.

The reliability of the data used to construct scoring keys is examined in the next section.

7. Reliability of Expert Judgements

The consistency among expert individual categorizations and weights of options was determined by a one-way analysis of variance with repeated measures. The results of these analyses for CPMPs 1-4 are presented in Table 6.103.

Reliabilities for the individual judgements and computer selections are reported for both one and n experts. The reliabilities for individual judgements range from 0.47 to 0.60 for r_1 (*i.e.*, the estimated reliability for one expert) and from 0.94 to 0.96 for r_n (*i.e.*, the estimated reliability of the mean rating of n experts). From this analysis it was concluded that the estimated reliability of individual mean judgements (*i.e.*, r_n) on CPMPs 1-4 was high.

The reliability estimates for computer selections range from 0.17 to 0.54 for r_1 and 0.67 to 0.92 for r_n. From this data it was concluded that the estimated reliability of the experts' computer selections was high on CPMPs 1, 2, and 4 (*i.e.*, 0.90 to 0.92) but low on CPMP 3 (*i.e.*, 0.67).

Table 6.103

Reliabilities of Individual Judgements
on Computer Selections on CPMPs 1-4

Individual Judgement	0.60*/0.94** (n = 16)	0.59/0.96 (n = 11)	0.47/0.94 (n = 16)	0.60/0.95 n = 10)
Computer Selections	0.48/0.91 (n = 11)	0.54/0.92 (n = 10)	0.17/0.67 (n = 10)	0.37/0.90 (n = 16)

$^{*}r_1$ = estimated reliability for one judge
$^{**}r_n$ = estimated reliability of the mean ratings of n judges

The estimates of reliability for examinee responses to homogeneous options is presented in the next section.

8. Estimates of Reliability for Homogeneous Options Within CPMPs

Cronbach's coefficient (alpha) and Lord's maximum estimate were calculated on homogeneous options within each CPMP for the proficiency, error of commission and error of omission scores. There were 36 alpha estimates for each CPMP (*i.e.*, 12 scoring procedures X 3 scores). These estimates are presented in Tables 6.104 to 6.107. In addition, the maximum latent root of the matrix of correlation coefficients among options is presented. The latent root was used in estimating the maximum

Table 6.104

Reliability Estimates for Proficiency, Error of Commission and Error of Omission Scores Calculated Using Twelve Scoring Procedures on CPMP 1

	Proficiency			Error of Commission			Error of Omission		
	largest eigenvalue	maximum	observed	largest eigenvalue	maximum	observed	largest eigenvalue	maximum	observed
Author	12.00	0.93	0.44	21.41	0.97	0.43	16.64	0.96	0.48
GCS	12.00	0.93	0.52	20.92	0.97	0.44	20.92	0.97	0.56
GDS	35.96	0.98	0.54	20.92	0.97	0.41	17.44	0.97	0.53
ICS	12.00	0.93	0.49	18.84	0.97	0.43	17.03	0.97	0.53
IDS	12.00	0.93	0.54	21.30	0.97	0.45	17.57	0.97	0.53
ICM	12.00	0.93	0.56	16.44	0.97	0.35	14.85	0.96	0.55
IDM	12.00	0.93	0.55	21.93	0.97	0.47	17.08	0.97	0.55
CCS	10.00	0.92	0.45	12.03	0.97	0.35	15.18	0.96	0.47
CDS	30.92	0.98	0.32	23.02	0.97	0.34	16.00	0.96	0.49
CCM	11.00	0.92	0.48	15.68	0.97	0.33	15.41	0.96	0.50
CDM	12.00	0.93	0.38	22.68	0.97	0.39	15.41	0.96	0.51
Mc	12.00	0.93	0.50	20.92	0.97	0.44	17.26	0.97	0.53

Table 6.105

Reliability Estimates for Proficiency, Error of Commission and Error of Omission Scores Calculated Using Twelve Scoring Procedures on CPMP 2

	Proficiency			Error of Commission			Error of Omission		
	largest eigenvalue	maximum	observed	largest eigenvalue	maximum	observed	largest eigenvalue	maximum	observed
Author	21.30	0.96	0.25	27.23	0.99	0.33	12.96	0.95	0.39
GCS	20.57	0.96	0.37	26.50	0.99	0.42	12.52	0.95	0.40
GDS	20.57	0.96	0.40	26.50	0.99	0.45	12.52	0.95	0.31
ICS	18.95	0.96	0.38	13.29	0.98	0.38	13.33	0.95	0.42
IDS	26.55	0.97	0.38	20.48	0.98	0.20	14.82	0.95	0.43
ICM	19.18	0.96	0.38	13.69	0.98	0.05	13.48	0.95	0.43
IDM	30.11	0.98	0.37	17.10	0.98	0.15	15.81	0.95	0.43
CCS	4.76	0.80	0.36	23.19	1.00	0.67	11.49	0.94	0.39
CDS	19.22	0.96	0.31	33.64	0.99	0.43	11.22	0.94	0.45
CCM	4.76	0.80	0.36	24.12	0.99	0.46	11.52	0.94	0.43
CDM	19.00	0.96	0.36	24.12	0.99	0.46	11.52	0.94	0.43
Mc	18.18	0.96	0.38	26.50	0.99	0.40	11.49	0.94	0.41

Table 6.106

Reliability Estimates for Proficiency, Error of Commission and Error of Omission Scores Calculated Using Twelve Scoring Procedures on CPMP 3

	Proficiency			Error of Commission			Error of Omission		
	largest eigenvalue	maximum	observed	largest eigenvalue	maximum	observed	largest eigenvalue	maximum	observed
Author	7.06	0.91	-0.63	1.00	0.00	0.00	6.96	0.90	-0.93
GDS	6.74*	0.88	-0.23	4.86	0.85	-0.09	7.22	0.91	-0.14
GDS	6.56*	0.88	-0.03	4.85	0.85	-0.36	7.23	0.91	-0.28
ICS	4.30	0.80	-0.25	1.18	0.20	-0.42	6.33	0.89	0.21
IDS	8.34*	0.91	0.16	4.84	0.85	0.07	7.86	0.91	-0.02
ICM	4.80	0.83	0.00	1.04	0.06	-0.10	6.57	0.89	-0.08
IDM	8.47	0.91	0.08	4.70	0.85	0.29	7.83	0.91	-0.18
CCS	6.36	0.89	0.17	1.56	0.42	0.02	5.51	0.88	-0.13
CDS	7.48	0.89	-0.20	4.89	0.85	-0.26	7.14	0.90	-0.10
CCM	6.73	0.90	0.03	1.34	0.37	0.01	6.53	0.89	-0.08
CDM	7.43	0.89	-0.14	3.76	0.64	-0.08	7.34	0.91	-0.15
Mc	7.12*	0.89	-0.33	4.86	0.85	-0.12	7.34	0.91	-0.34

*no convergenceon 10 trials

Table 6.107

Reliability Estimates for Proficiency, Error of Commission and Error of Omission Scores Calculated Using Twelve Scoring Procedures on CPMP 4

	Proficiency			Error of Commission			Error of Omission		
	largest eigenvalue	maximum	observed	largest eigenvalue	maximum	observed	largest eigenvalue	maximum	observed
Author	14.43	0.95	0.57	9.35	0.94	-0.55	4.40	0.81	0.76
GCS	14.67	0.95	0.69	9.13	0.94	-0.32	4.60	0.81	0.76
GDS	14.64	0.95	0.63	9.13	0.94	-0.29	4.44	0.80	0.75
ICS	13.64	0.95	0.58	8.33	0.94	0.01	3.57	0.75	0.65
IDS	16.94	0.96	0.62	10.59	0.95	-0.05	4.18	0.79	0.67
ICM	13.10	0.95	0.57	8.08	0.94	0.01	2.95	0.69	0.62
IDM	17.15	0.96	0.64	10.59	0.95	-0.06	4.35	0.80	0.67
CCS	13.45	0.95	0.67	7.27	0.94	-0.05	3.70	0.76	0.68
CDS	18.71	0.97	0.46	13.93	0.96	-0.05	3.30*	0.72	0.52
CCM	11.77	0.94	0.48	7.27	0.94	-0.05	2.87	0.69	0.48
CDM	18.78	0.97	0.53	14.19	0.96	0.00	3.12*	0.71	0.58
Mc	14.37	0.95	0.56	9.13	0.94	-0.36	3.80*	0.76	0.62

*no convergence on 10 trials

alpha coefficient. Each score of proficiency, error of commission and omission for each CPMP will be discussed in turn.

A. Reliability Estimates for CPMP 1:

There is little variation in alpha maximum coefficients among scoring procedures: proficiency from 0.92 to 0.98, error of commission from 0.97 to 0.98, and error of omission from 0.96 to 0 97. Greater variation was observed among the observed alphas: proficiency from 0.32 to 0.56, error of commission from 0.33 to 0.47 and error of omission from 0.47 to 0.56 (see Table 6.104).

B. Reliability Estimates for CPMP 2:

Basically, the same results were found in CPMP 2 as in CPMP 1. The maximum alphas varied from 0.80 to 0.96, from 0.98 to 1.00 and from 0.94 to 0.95 respectively for proficiency, error of commission and omission scores. Greater variation was observed among the observed alphas: proficiency from 0.25 to 0.40, error of commission from 0.05 to 0.46 and error of omission from 0.31 to 0.46 (see Table 6.105).

C. Reliability Estimates for CPMP 3:

Table 6.106 presents the reliability estimates for examinee scores on CPMP 3.[1] The most striking feature of this table is that the observed alpha coefficients vary from -0.93 to 0.29. Since the correlation coefficient is an estimate of the test reliability and since the reliability for a test composed of two parallel components is theoretically positive, the least estimate of internal consistency would be zero. It therefore must be concluded that the test was composed of two or more homogeneous but independent sub-tests or that the homogeneous options had no internal consistency.

---

[1] In order to reduce computer costs for calculating the largest eigenvalue, the maximum number of iterations for convergence was set to 10. The * indicates the eigenvalues calculated after 10 iterations.

D. Reliability Estimates for CPMP 4

The maximum alphas varied from 0.94 to 0.97 for proficiency, from 0.94 to 0.96 for error of commission and from 0.69 to 0.81 for error of omission (see Table 6.107). The observed alphas however respectively varied from 0.46 to 0.69, -0.55 to 0.01 and 0.48 to 0.76. The best estimate of observed alphas for error of commission scores was zero. This result is best explained in light of the structure of CPMP 4. In this test, the examinee had seven opportunities to refer the patient to an obstetrician. Forty-six examinees elected to do so and therefore did not reach the end of the problem. Most of the negative scores in the test were assigned to those options which referred the patient to the obstetrician. Since this option was selected once and the encounter terminated, there was no consistency in responses among the negative options.

From the above results it was observed that variations occurred in reliability estimates across scoring procedures within the four CPMPs. From this observation it was concluded that the consistency of responses on homogeneous options could be effected by different scoring procedures.

9. Validity Measure

The mean scores of expert problem-solvers were compared to a minimal performance level (MPL) to determine whether all problem solvers were "experts". Table 6.108 summarizes the number of expert problem-solvers with scores above the MPL. The table indicates that all or almost all experts achieved the MPLs of 75% and 70% respectively on CPMPs 1 and 2, that as many as five out of ten did not achieve the MPL of 60% on CPMP 3, and as many as three out of sixteen did not achieve the MPL of 70% on CPMP 4. From these observations it appeared that groups A and C who respectively solved CPMPs 1 and 2, were probably "experts", but perhaps not all members of groups A and B, who respectively solved CPMPs 3 and 4, were "experts". Five out of ten members of group A were branched to the mismanagement side of the problem because they had not recognized that a gynecologist would perform a laporotomy on the patient after being referred. In CPMP 4 three problem-solvers referred the patient to an obstetrician nearly half-way through the clinical encounter. Although the referral may have been an appropriate decision for them (_i_._e_., one was a surgeon) it was judged to be below the acceptable standard set for the examinee. Therefore, three of the problem-solvers were not "experts" with respect to this problem.

Table 6.108

Number of Expert Problem-Solvers
With Scores Above the Minimal Pass Level (MPL)

CPMP	Author	GCS	GDS	ICS	IDS	ICM	IDM	CCS	CDS	CCM	CDM	Mc
1 (N=11)	11	11	11	11	11	11	11	11	11	11	11	11
2 (N=10)	10	9	10	10	10	10	10	10	10	10	10	10
3 (N=10)	8	6	8	6	5	7	5	6	10	10	10	8
4 (N=16)	16	14	14	13	14	13	14	13	16	13	15	14

On the basis of the above observations, it was concluded that groups A and C, who respectively solved CPMPs 1 and 2, were "experts"; five out of ten members of group A, who solved CPMP 3; and thirteen out of sixteen members of group B, who solved CPMP 4, were experts.

A summary of the conclusions, implications, limitations and recommendations for further research are presented in the next chapter.

CHAPTER VII

SUMMARY, IMPLICATIONS, LIMITATIONS AND RECOMMENDATIONS

1. Summary

This study was designed to determine the possible effects of various scoring procedures upon examinee CPMP scores. Some light was shed upon the matter, but the study did not and could not identify which one of the twelve scoring procedures investigated was optimal. The study was able to identify that different scoring procedures do have varying effects upon examinee scores and that these effects are in some cases dependent upon the CPMP used. More specifically, it was found that scoring procedures could alter the:

1) shape of the distribution of scores
2) score variance
3) validity of the trait or behavior being measured
4) score mean
5) examinee satisfactory/unsatisfactory status
6) test/retest reliability, and
7) rank ordering of examinees.

Before each of the above aspects is summarized, some comments are in order regarding the characteristics of the scoring keys and the group of "expert" judges employed to categorize and weight CPMP options.

Firstly, it was observed that weightings for the same option could vary greatly between scoring procedures (i.e., from an indispensibly positive to an unforgivable negative). Scoring procedures also varied in the number of positive and negative options identified for the same CPMP. These variations in option weights and categorizations were partly due to the specific characteristics of each scoring procedure and partly to the consistency or inconsistency of weightings given options by the "experts". The following is a summary of the effects of the above variations upon examinee scores.

A. Shape of the Distribution of Scores

It had been observed that both the kurtosis and skewness of scores varied among scoring procedures and that this variation was dependent upon the CPMP and the score (i.e., proficiency, error of omission and efficiency). Since there was no trend produced by the scoring procedures it could only be concluded that scoring procedures could, in general, alter the distribution of scores.

B. Score Variance

A statistical comparison of the variance of scores revealed many significant differences among scores calculated by the various scoring procedures. Once again, as no trend was produced by the different scoring procedures, it could only be concluded that, in general, scoring procedures could significantly alter the variance of scores.

C. Trait or Behavior Measured

A principal components factor analysis was undertaken to determine what specific traits were actually measured by each CPMP and whether these measurements were altered by the scoring procedures. When the scores generated by each of the scoring procedures for each CPMP were factor analyzed, it was observed that only one CPMP, but several scoring procedures loaded on a factor. Since no two CPMPs loaded on the same factor, it was concluded that clinical performance, as measured by the computer simulations, was case specific. However, not all scoring procedures loaded on the same factor. The scoring procedures were observed to fall on different factors depending upon the CPMP, the method of categorization and the score analyzed. It was therefore concluded that the scoring procedures could produce measures of different behaviors or different measures of the same behavior.

In order to further study this alteration, the scores of each CPMP were factor analyzed. This lead to the observation that the number of factors varied over CPMPs and scores analyzed. The number of factors appeared to be most highly related to the structural complexity of the CPMP and, to a lesser degree, the method of categorization employed.

A trend of loadings was noted among the methods of categorizing options. The author and group methods loaded together on the same factor; the computer method by itself;

and, the individual method either by itself, with the author and group methods, or with the computer method. From these observations it was concluded that a similar behavior is measured by the author and group methods which is unlike that behavior measured by the computer method and which may or may not be similar to the behavior measured by the individual method. It was also observed that the weights (constant/differential) and number of keys (single/multiple) had no effect upon the factor loadings.

D. Mean Scores

A multivariate analysis was undertaken to determine the effect of scoring keys upon examinee mean scores. No consistent pattern was observed among the CPMP's but trends did exist.

Firstly, it was observed that the author's mean proficiency score tended to be the lowest and the computer's the highest. In addition, the computer's error of omission score tended to be lower than that of other scoring procedures. These results occurred due to the relatively closer match between examinee selections and options weights in the computer procedure.

Secondly, the individual scoring procedure tended to yield the highest efficiency scores. This result occurred due to the relatively greater number of positive options in the individual scoring procedure.

Thirdly, it was observed that the McLaughlin and GDS mean scores were statistically equal on CPMPs 1-4 even though these procedures employed different methods in assigning differential weights.

Fourthly, it was observed that the mean proficiency scores were higher for scoring procedures with differential weights than for those with constant weights, while the error of omission scores were lower. These results were explained by examining the equation for the efficiency percentage.

Lastly, it was observed that scores did not differ relative to the key employed (single or multiple). This was attributed to the inability of methods employed to separate small numbers of experts into two or more homogeneous groups.

E. Examinee Satisfactory (Pass)/Unsatisfactory (Fail) Status

The number of examinees declared satisfactory/unsatisfactory varied greatly among scoring procedures. This observation was particularly pronounced in CPMPs 2, 3 and 4. When CPMP 2 was further analyzed to determine the extent of variation, it was found that status changed on as few as zero and as many as 53 out of 111 examinees.

F. Test/Retest Reliability

Cronbach's alpha coefficient and Lord's maximum estimates were calculated for the proficiency, error or commission and omission scores based upon homogeneous options scored by each of the twelve scoring procedures. It was observed that the

Lord's maximum estimates approached 1.0 but the observed alphas varied from 0.0 to 0.76 depending upon the CPMP and the score analyzed. Since there was no pattern of change among the scoring procedures, it could only be concluded that indices reflecting the consistency of responses on homogeneous options could be affected by different scoring procedures.

G. Rank Ordering of Examinee Scores

The rank ordering of examinee proficiency scores was calculated using Spearman's rho. A statistical test of $\rho_{xy} - \rho_{xz}$ for dependent samples was used to determine significant differences among the examinee rankings. It was observed that significant differences occurred among rankings, but no pattern or trend was evident over CPMPs. It was therefore concluded that scoring procedures do significantly alter examinee rank orderings but that specific causal relationships could not be defined.

H. Conclusions

On the basis of this investigation, the following conclusions were reached regarding the effects of different scoring procedures upon examinee CPMP scores:

1) the weightings (categorization + weight) can vary greatly over scoring procedures. These variations are partly due to the specific characteristics of each scoring key and partly to the consistency or inconsistency of weightings given

options by experts,

2) the distribution of examinee scores (*i.e.*, skewness, kurtosis and variance) can change with different scoring procedures

3) clinical performance on CPMPs was problem specific,

4) an examinee's clinical decision-making score is primarily dependent upon the content of the CPMP but the scoring procedure (*i.e.*, method of categorization) can alter the behavior that is measured,

5) the author and group methods of categorization measure similar behaviors which can differ from those behaviors measured by the computer and individual methods,

6) the mean proficiency score for the computer method is higher than that for other scoring procedures while its error of omission score is lower,

7) scoring procedures with differential weights yield higher examinee mean scores than those with constant weights,

8) scoring procedures using individual judgements yield the largest number of positive (+) options which in turn result in the largest examinee mean efficiency scores,

9) scores generated by the McLaughlin and GDS scoring procedures are equivalent,

10) there is no difference between examinee mean scores generated using single and multiple keys,

11) the satisfactory/unsatisfactory status of examinees can vary depending upon the scoring procedure employed,

12) the method of categorizing options could alter the measure of internal consistency,

13) scoring procedures alter the rank ordering of examinees, and

14) the more complex the structure of the simulation, the more difficult it is to develop a valid scoring key.

In summary, it was concluded that scoring procedures can be an added source of variability in examinee CPMP scores.

2. Implications

This investigation was of particular importance to the author because it revealed how little we know about scoring simulations which purports to measure clinical problem-solving skills. Variations in scoring procedures can alter examinee scores. It has been shown that both the method of categorization (author, group, individual or computer) and the type of weight assigned (constant or differential) can affect examinee scores.

The scoring keys investigated in this study employed the group, individual or computer methods of categorization. In the author and group methods, options were categorized by experts having prior knowledge of the correct solution to the problem. It would seem that keys generated from the group and author categorizations would reflect a different mode of behavior than that employed by the examinee. The examinee, having no prior knowledge of the solution, selects options using problem-solving behavior while the expert categorizes options with knowledge of the correct solution. It is therefore suggested that the scores generated by these scoring keys would reflect the measurement of these two different types of behavior. Perhaps the practice of generating scoring keys while possessing prior knowledge of the correct solution is one reason why scores generated by the author and group consensus methods frequently loaded on the same factor.

On the other hand, options that are categorized and weighted while problem-solving are closer to the task faced by the examinee. Thus, an examinee score generated by scoring keys which employ computer categorizations would reflect the selections made by both the expert and the examinee while in the problem-solving mode of behavior. Perhaps it was for this reason that the scores generated by the computer procedure tended to load on separate factors and have higher means.

From the above observations, it is suggested that a more optimal method for categorization of options would be to employ:

1) both group consensus and computer performance methods: group consensus being used to refine the weights derived from computer performance, or

2) group consensus to categorize and weight options while problem-solving.

Use of either of the above methods would help to ensure that scores reflect the problem-solving process as well as the degree to which each task is correctly completed.

It has also been observed that the type of weight used by the scoring procedure can affect examinee scores. Based upon the results of this study, it is suggested that options which are of varying degrees of importance in the resolution of the patient's problem be given differential weights. The weights would then reflect the contribution of the option in resolving the patient's problem.

The above discussion focuses on the alterations in examinee performance resulting from variations in the twelve scoring procedures. However, it is important to remember that all of these scoring procedures are based upon an additive model which could also invalidate examinee scores since it may not reflect the degree to which a task is correctly completed (i.e., the additive affect of several choices may be much greater than their sum). For example, if an examinee gave five forgivable treatments (e.g., drugs) to a patient, individually, they may have no serious repercussions, but collectively, they may be deadly. That is, the interaction of the drugs may have a multiplicative rather than an additive effect. Further investigation is required to determine the inadequacies of the additive model.

The twelve scoring procedures investigated in this study have another common characteristic. The attempt to summarize the complex process of clinical problem-solving in a single score resulted in a great deal of information being lost. Rather than using a single score to summarize examinee performance, it would seem more appropriate to think of clinical problem-solving as a profile of abilities observed in an appropriate sample of clinical cases. This is particularly true by the scoring procedure can affect examinee scores. Based upon the results of this study, it is suggested that options which are of varying degrees of importance in the resolution of the patient's problem be given differential weights. The weights would then reflect the contribution of the option in resolving the patient's problem.

The above discussion focuses on the alterations in examinee performance resulting from variations in the twelve scoring procedures. However, it is important to remember that all of these scoring procedures are based upon an additive model which could also invalidate examinee scores since it may not reflect the degree to which a task is correctly completed (*i.e.*, the additive effect of several choices may be much greater than their sum). For example, if an examinee gave five forgivable treatments (*e.g.*, drugs) to a patient, individually, they may have no serious repercussions, but collectively, they may be deadly. Further investigation is required to determine the prevalence of this problem.

The twelve scoring procedures investigated in this study have another common characteristic. The attempt to summarize the complex process of clinical problem-solving in a single score resulted in a great deal of information being lost. Rather than using a single score to summarize examinee performance, it would seem more appropriate to think of clinical problem-solving as a profile of abilities observed in an appropriate sample of clinical cases. This is particularly true in medical schools where more emphasis is placed on the clinical process of problem-solving and less on the outcome itself. Thus a student's profile may indicate strengths and weaknesses in:

1) various components of problem-solving (*e.g.*, hypothesis generation, data gathering, data interpretation, data utilization, and hypothesis refinement), and

2) particular types of cases (*e.g.*, cardiac problems, emergency problems, obstetrical problems in young females, etc.). Such profiles should be based upon a sample of many cases. In all profiles, the generated score should reflect the degree to which each task and/or process is correctly completed.

Emphasis upon clinical problem-solving processes may however be inappropriate for licensing agents charged with the responsibility of assessing clinical competence. Instead, their emphasis should be directed toward the proper use of available resources for optimal patient management.

However, both medical schools and licensing agents should be concerned with the reliability and validity of simulations. Therefore, they should examine the effects that variations in the structure of the simulation may have upon examinee performance. For example, variation in performance could occur due to the:

1) response mode - selection or open response
2) method of responding - rubout or typed
3) number of pathways - one (linear) to unlimited
4) time - static or dynamic

5) presentation of information - verbal description, visual and/or audio.

With respect to variations in pathways, the results revealed that particular caution should be exercised in the development of scoring keys for branching problems. The judge(s) must thoroughly understand the problem and assign weights which reflect the merit of each pathway in the problem's resolution. In order for this to occur, it would appear that judge(s) should know the solution to the patient's problem. However, as discussed earlier, prior knowledge of the correct solution to a problem may have an unfavourable effect upon the scoring key. It is therefore suggested that expert problem-solvers be used to categorize options and group consensus be used to determine the relative merits of each pathway. By using this method, one can insure that weightings more closely match the task of the problem-solver and that scores reflect the degree to which the task was correctly completed.

From the above discussions, it is evident that examinee CPMP scores are affected by a wide range or variations. With all of these sources of variation, one can only be impressed with the complexity and elusiveness of clinical problem-solving and by the promise that simulations hold in furthering our understanding of this process.

In summary, it is recommended that:

1) a profile of student clinical performance be generated,

2) licensing agents emphasize proper use of available resources for optimal patient management,

3) greater attention be paid to the potential use of expert problem-solvers' performance in the establishment of scoring keys,

4) differential weights be employed which reflect the option's contribution in resolving the patient's problem, and

5) continued efforts be made to understand and measure clinical problem-solving.

3. Limitations

It is difficult to generalize the results of this study beyond the fourth year medical students and medical problems used in this study since neither students nor medical problems were randomly sampled. It is also important to keep in mind that the categorization of options in CPMPs is a judgemental process. Thus, the usefulness of the types of scoring procedures generated within this study is dependent upon the quality of the judgements and the procedures used to reduce or eliminate descrepancies within these judgements. It was also observed that not all "experts" were experts, and to this extent, the results are also limited.

4. Recommendations

This investigation has provided insights into the effects of scoring procedures upon examinee scores. Based upon the results and insights gained within this study, it is recommended that the following be investigated:

1) the processes underlying clinical problem-solving,

2) the effect of the nature of the problem upon clinical problem-solving performance,

3) the effect of the structure of the simulation upon clinical problem-solving performance,

4) the merits of categorizing and assigning weights to options while problem-solving,

5) possible procedures for determining examinee scores other than the additive model, and

6) the consistent but different perceptions among problem-solvers.

LIST OF REFERENCES

Allal, L.K. Training of Medical Students in a Problem-Solving Skill: The Generation of Diagnostic Problem Formulations. Unpublished doctoral dissertation, Michigan State University, 1973.

Barrows, H.S., & Bennett, K. The Diagnostic (Problem Solving) Skill of the Neurologist. Arch. Neurol, 1972, 26, 273-277.

Berner, E.S., Bligh, T.J., & Guerin, R.O. An Indication for a Process Dimension in Medical Problem Solving. Medical Education, 1977, 11(5), 324-328.

Berner, E.S., Hamilton, L.A., & Best, W.R. A New Approach to Evaluating Problem-Solving in Medical Students. Journal of Medical Education, 1974, 49, 666-672.

Bloom, B.S. Taxonomy of Educational Objectives - The Classification of Educational Goals - Handbook I: Cognitive Domain. New York: McKay, 1956

Babula, J.A., & Page, G.G. The Construction of Written Simulations. Research and Evaluation Section Center for Educational Development, University of Illinois, College of Medicine, Chicago, Illinois, 1973.

Cattell, R.B. Validity and Reliability: A proposed more basic set of concepts. Journal of Educational Psychology, 1964, 55(1), 1-22.

Cronbach, L.J. Coefficient Alpha and Internal Structure of Test. Psychometrika, 1951, 16, 274-334.

Cronbach, L.J. & Rajaratnam, N. Theory of Generalizability: A literalization of reliability theory. British Journal of Statistical Psychology, 1963, 16, 137-163.

Donnelly, M.B., Gallagher, R.E., Hess, J.W., & Hogan, M.J. The Dimensionality of Measures Derived from Complex Clinical Simulations. Proceedings, Conference on Research in Medical Education, Association of American Medical Colleges, 1974.

Elstein, A.S., Kagan, N., Schulman, L.S., Jason, H., & Loure, M.J. Methods and Theory in the Study of Medical Inquiry. Journal of Medical Education, 1972, 5, 48-52.

Elstein, A.S., Shulman, L.S. & Sprafka, S.A. Medical Problem Solving. Cambridge: Harvard University Press, 1978.

Evaluation Unit. Simulation Technique in the Evaluation of Clinical Judgement. Center for the Study of Medical Education, University of Illinois, College of Medicine, 1967, 7.

Friedman, R.B. A Computer Program for Simulating the Patient-Physician Encounter. Journal of Medical Education, 1973, 48, 92-97.

Glass, G.V. & Stanley, J.C. Statistical Methods in Education and Psychology. Englewood Cliffs, New Jersey: Prentice-Hall Inc., 1970, 313-314.

Goldberg, L.R. Man Versus Model of Man: A Rationale, Plus Some Evidence for a Method of Improving on Clinical Inferences. Psychological Bulletin, 1970, 73, 422-432.

Hammond, K.R., Hursch, C.J., & Todd, F.J. Analyzing the Components of Clinical Inference. Psychological Review, 1964, 71, 438-456.

Hampton, J.R., Harrison, M.J.G., Mitchell, J.R.A., Pritchard, J.S., & Seymour, C. Relative Contributions of History-Taking, Physical Examination and Laboratory Investigations to Diagnosis and Management of Medical Outpatients. British Journal of Medical Education, 1975, 2, 486-489.

Harless, W.G., Drennon, G.G., Marxer, J.J., Rott, J.A., & Miller, G.E. CASE: A Computer-Aided Simulation of the Clinical Encounter. Journal of Medical Education, 1971, 46, 443-448.

Helfer, R.E. & Slater, C.H. Measuring the Process of Solving Clinical Diagnostic Problems. British Journal of Medical Education, 1971, 5, 48-52.

Hoffman, D. The Paramorphic Representation of Clinical Judgement. Psychological Bulletin, 1960, 57, 116-131.

Hubbard, J.P. Measuring Medical Education. Philadelphia: Lea and Febiger, 1971.

Juul, D.H., Noe, M.J. & Nerenberg, R.L. *A Factor Analytic Study of Branching Patient Management Problems*. Paper presented at the annual meeting of the National Council of Measurement in Education, New York City, April, 1977.

Kagan, N., Elstein, A.S., & Jason, H. *A Theory of Medical Inquiry*. Paper read before the Ninth Annual Conference on Research in Medical Education, Los Angeles, 1970.

Kleinmuntz, B. *Formal Representation of Human Judgement*. New York: John Wiley & Son, 1968.

Leaper, D.J., Gill, P.W., Staniland, J.R., Horrocks, J.C., & DeDombal, F.T. Clinical Diagnostic Process: An Analysis. *British Journal of Medical Education*, 1973, *3*, 569-574.

Lord, F.M. Some Relations Between Guttman's Principal Components of Scale Analysis and Other Psychometric Theory. *Psychometrika*, 1958, *23*, 291-296.

Lord, F.M. & Novick, M.R. *Statistical Theories of Mental Test Scores*. Don Mills, Ontario: Addison Wesley, 1968.

McGuire, C.H. Simulation Technique in the Teaching and Testing of Problem Solving Skills. *Journal of Research in Science Teaching*, 1976, *13*(2), 89.

McGuire, C.H. & Babbott, D. Simulation Techniques in the Measurement of Problem Solving Skills. *Journal of Educational Measurement*, 1967, *4*, 1-10.

McGuire, C.H. & Soloman, L.M. (Eds.) *Clinical Simulations: Selected Problems in Patient Management*. New York: Appleton, Century, Crofts, 1971.

Miller, G.A. The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information. *Psychological Review*, 1956, *63*, 81-97.

Rumoldi, H.J.A. The Test of Diagnostic Skills. *Journal of Medical Education*, 1961, *36*, 73-79.

Rumoldi, H.J.A. Rationale and Applications of the Test of Diagnostic Skills. *Journal of Medical Education*, 1963, *38*, 364-368.

Shulman, L.S. The Psychology of School Subjects: A Premature Obituary. *General Research and Science Teaching*, 1974, *11*, 319-339.

Simon, H.A. The Sciences of the Artificial. Cambridge, Massachusetts: M.I.T. Presentation, 1969.

Simon, H.A., & Newell, A. Human Problem Solving: The State of the Theory in 1970. American Psychology, 1971, 26, 145-159.

Skakun, E.N. The Dimensionality of Linear Patient Management Problems. Proceedings of the Seventeenth Annual Conference on Research in Medical Education, New Orleans, Louisiana, October, 1978, 63-67.

Skakun, E.N. The Use and Construction of Patient Management Problems in Evaluating Medical Students. A paper presented at the 10th Annual Canadian Cardiovascular Society Symposium on Teaching and Evaluation, Edmonton, Alberta, June 20-21, 1975.

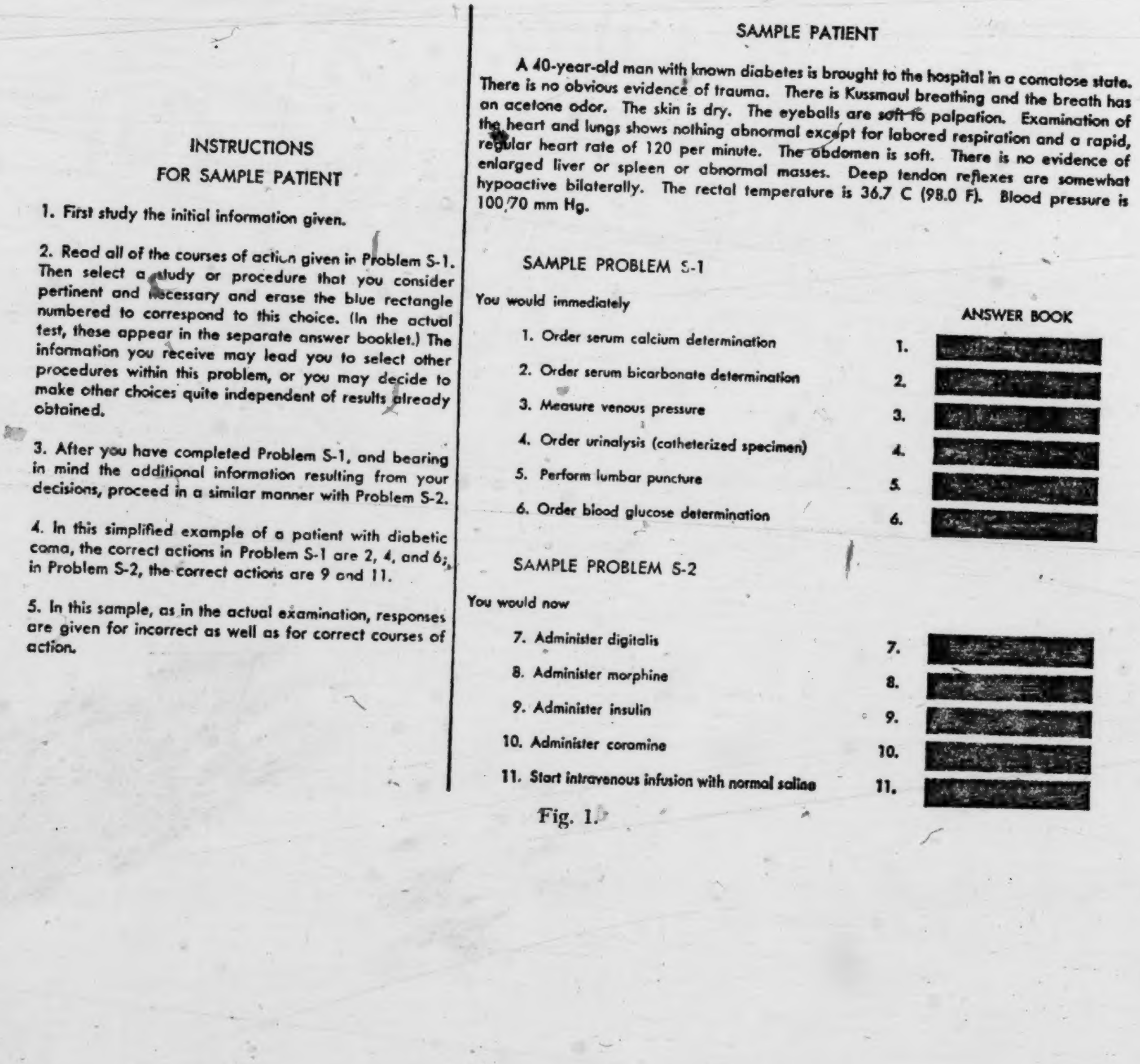
Taylor, T.R., Aitchison, J., Parker, L.S., & Moore, M.F. Individual Differences in Selecting Patients for Regular Dialysis. British Journal of Medical Education, 1975, 2, 380-381.

Wason, P.C. & Johnson-Laird, P.N. Psychology of Reasoning: Structure and Content. Cambridge, Massachusetts: Harvard University Press, 1972.

Weed, L. Medical Records, Medical Education and Patient Care. Cleveland, Ohio: (The Press of) Case Western Reserve University, 1970.

APPENDIX A

Sample of Linear Simulated Patient Encounter



INSTRUCTIONS FOR SAMPLE PATIENT

1. First study the initial information given.

2. Read all of the courses of action given in Problem S-1. Then select a study or procedure that you consider pertinent and necessary and erase the blue rectangle numbered to correspond to this choice. (In the actual test, these appear in the separate answer booklet.) The information you receive may lead you to select other procedures within this problem, or you may decide to make other choices quite independent of results already obtained.

3. After you have completed Problem S-1, and bearing in mind the additional information resulting from your decisions, proceed in a similar manner with Problem S-2.

4. In this simplified example of a patient with diabetic coma, the correct actions in Problem S-1 are 2, 4, and 6; in Problem S-2, the correct actions are 9 and 11.

5. In this sample, as in the actual examination, responses are given for incorrect as well as for correct courses of action.

SAMPLE PATIENT

A 40-year-old man with known diabetes is brought to the hospital in a comatose state. There is no obvious evidence of trauma. There is Kussmaul breathing and the breath has an acetone odor. The skin is dry. The eyeballs are soft to palpation. Examination of the heart and lungs shows nothing abnormal except for labored respiration and a rapid, regular heart rate of 120 per minute. The abdomen is soft. There is no evidence of enlarged liver or spleen or abnormal masses. Deep tendon reflexes are somewhat hypoactive bilaterally. The rectal temperature is 36.7 C (98.0 F). Blood pressure is 100/70 mm Hg.

SAMPLE PROBLEM S-1

You would immediately

	ANSWER BOOK
1. Order serum calcium determination	1.
2. Order serum bicarbonate determination	2.
3. Measure venous pressure	3.
4. Order urinalysis (catheterized specimen)	4.
5. Perform lumbar puncture	5.
6. Order blood glucose determination	6.

SAMPLE PROBLEM S-2

You would now

7. Administer digitalis	7.
8. Administer morphine	8.
9. Administer insulin	9.
10. Administer coramine	10.
11. Start intravenous infusion with normal saline	11.

Fig. 1.

APPENDIX B.1

MENINGITIS MANAGEMENT SECTION

THE SCORING IS BASED ON THE PERFORMANCE OF 11 PEDIATRICIANS

SECTION	ITEM	SCORE	NO. OUT OF 11 SPECIALISTS WHO DID TAKE + ITEMS	NO. OUT OF 11 SPECIALISTS WHO DID NOT TAKE - ITEMS	WEIGHTING +%	WEIGHTING -%
A IMMEDIATE EVALUATION	1	+	3		1.9	
	2	+	1		0.6	
	3	+	7		4.4	
	4	-		11		3.8
	5	-		6		2.1
	6	+	9		5.6	
	7	-		7		2.4
	8	-		7		2.4
	9	-		9		3.1
	10	+	11		6.9	
	11	+	11		6.9	
	12	-		7		2.4
	13	-		10		3.5
B INITIAL TREATMENT	1	+	7		4.4	
	2	-		10		3.5
	3	-		11		3.8
	4	+	10		6.3	
	5	-		11		3.8
	6	-		9		3.1
	7	-		8		2.8
	8	-		6		2.1
	9	+	4		2.5	
C YOU WOULD NOW ORDER	1	-		7		2.4
	2	0				
	3	+	4		2.5	
	4	-		9		3.1
	5	-		11		3.8
	6	+	6		3.8	
	7	+	2		1.3	
	8	-		11		3.8
	9	+	11		6.9	
D AFTER CANCELLATION OF PREVIOUS ORDERS YOU WOULD NOW ORDER	1	-		9		3.1
	2	-		11		3.8
	3	+	10		6.3	
	4	-		11		3.8
	5.	-		11		3.8
	6	-		11		3.8
	7	-		8		2.8
	8	-		11		3.8
	9	-		7		2.4
	10	-		9		3.1
	11	+	1		0.6	
	12	-		7		2.4

APPENDIX B.2

SECTION	ITEM	SCORE	NO. OUT OF 11 SPECIALISTS WHO DID TAKE + ITEMS	NO. OUT OF 11 SPECIALISTS WHO DID NOT TAKE - ITEMS	WEIGHTING +%	WEIGHTING -%
E AFTER 3 DAYS YOU WOULD ORDER	1	+	11		6.8	
	2	+	11		6.8	
	3	+	9		5.6	
	4	+	8		5.0	
	5	+	3		1.9	
	6	-		11		3.8
F YOUR PLAN OF MANAGEMENT WOULD NOW INCLUDE	1	-		10		3.5
	2	-		10		3.5
	3	+	8		5.0	
	4	-		10		3.5
	5	+	8		5.0	
	6	0				
	7	0				
	8	+	5		3.1	
TOTAL					100.0%	100.0%

APPENDIX C

SCORING FORMULAS FOR SIMULATION EXERCICES

<table>
<tr><th>Score</th><th>Formulation</th><th>Example</th></tr>
<tr>
<td>Proficiency (%)</td>
<td>The sum of (+) and (-) points for options chosen, divided by the maximum possible score, converted to percent.
$P = \frac{\Sigma[(+) + (-)]}{\text{Max. Score}} \times 100$</td>
<td>Candidate X made the following choices on a PMP where 90 was maximum score.
<table>
<tr><th>No. of Choices</th><th>Weight</th><th>Sum</th></tr>
<tr><td>3</td><td>16</td><td>48</td></tr>
<tr><td>2</td><td>8</td><td>16</td></tr>
<tr><td>4</td><td>2</td><td>8</td></tr>
<tr><td>2</td><td>0</td><td>0</td></tr>
<tr><td>2</td><td>-1</td><td>-2</td></tr>
<tr><td>2</td><td>-4</td><td>-8</td></tr>
</table>
$P = \frac{[72] + [-10]}{[90]} \times 100$
P= 68.8 or 69%</td>
</tr>
<tr>
<td>Errors of Omission (%)</td>
<td>100% minus [The sum of the positive points chosen, divided by maximum possible score, converted to per cent.]
$E.O.\% = (100) - \frac{\Sigma(+)}{\text{Max. Score}} \times 100$</td>
<td>For the above candidate:
$E.O. = 100 - \frac{72}{90} \times 100$
E.O. = 20%</td>
</tr>
<tr>
<td>Errors of Commission (%)</td>
<td>The sum of the negative points chosen, divided by the maximum possible score, converted to per cent.
$E.C.\% = \frac{\Sigma(-)}{\text{Max. Score}} \times 100$</td>
<td>For the above candidate:
$E.C.\% = \frac{10}{90} \times 100$
E.C.% = 11%
NOTE:
100% - [E.O.% + E.C.%] = P%
100% - [20% + 11%] = 69%</td>
</tr>
<tr>
<td>Efficiency (%)</td>
<td>The number of positively weighted choices made, divided by the total number of choices made, converted to per cent.
$E\% = \frac{\text{No. of (+) choices}}{\text{No. of all choices}} \times 100$</td>
<td>For the candidate:
9 choices were (+)
2 choices were 0
4 choices were (-)
15
$E = \frac{9}{15} \times 100$
E = 60%</td>
</tr>
</table>

APPENDIX D (linear)

INSTRUCTIONS

1. Read the patient management problem and become familiar with the final diagnosis and the various options offered.

2. Categorize each decision into one of the five categories:

++ (+2) Category:	Choices which are CLEARLY INDICATED and IMPORTANT in the care of THIS patient at THIS stage in the workup or management;
+ (+1) Category:	Choices which are CLEARLY INDICATED but of a more ROUTINE nature, i.e., should be selected but are not of special significance in the care of THIS patient at THIS stage;
0 Category:	Choices which are OPTIONAL, i.e., the probability that they will be helpful for THIS patient at THIS stage is fairly remote or quite debatable;
- (-1) Category:	Choices which are clearly NOT INDICATED though NOT HARMFUL in the management of THIS patient at THIS stage;
-- (-2) Category:	Choices which are clearly CONTRA-INDICATED (i.e., are definitely harmful or carry an unjustifiable high cost in terms of risk, pain or money) in the care of THIS patient at THIS stage.

3. Re-examine the options categorized as either ++ (+2) or -- (-2). Where appropriate these options should be further categorized as either +++ (+3), ++++ (4), +++++ (+5), or --- (-3), ---- (-4) ----- (-5), depending upon the perceived degree of their appropriateness or inappropriateness.

NOTE: Steps 2 and 3 may be carried out simultaneously.

APPENDIX E

(branching)

INSTRUCTIONS

1. Read the patient management problem and become familiar with the final diagnosis and the various options offered.

2. Using the flowchart outline the optimal route.

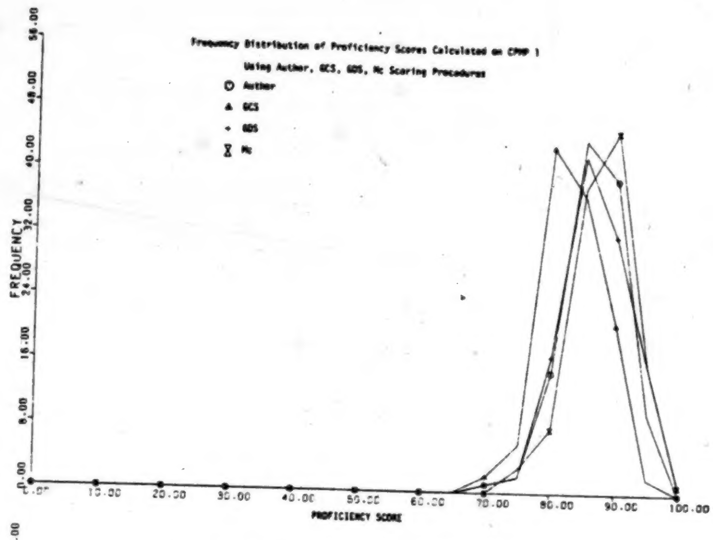
3. Categorize each decision into one of the following five categories:

++ (+2) Category:	Choices which are CLEARLY INDICATED and IMPORTANT in the care of THIS patient at THIS stage in the workup or management;
+ (+1) Category:	Choices which are CLEARLY INDICATED but of a more ROUTINE nature, i.e., should be selected but are not of special significance in the care of THIS patient at THIS stage;
0 Category:	Choices which are OPTIONAL, i.e., the probability that they will be helpful for THIS patient at THIS stage is fairly remote or quite debatable;
- (-1) Category:	Choices which are clearly NOT INDICATED though NOT HARMFUL in the management of THIS patient at THIS stage;
-- (-2) Category:	Choices which are clearly CONTRA-INDICATED (i.e., are definitely harmful or carry an unjustifiable high cost in terms of risk, pain or money) in the care of THIS patient at THIS stage.

4. Re-examine the options categorized as either ++ (+2) or -- (-2). Where appropriate these options should be further categorized as either +++ (+3), ++++ (4), +++++ (+5), or --- (-3), ---- (-4) ----- (-5), depending upon the perceived degree of their appropriateness or inappropriateness.

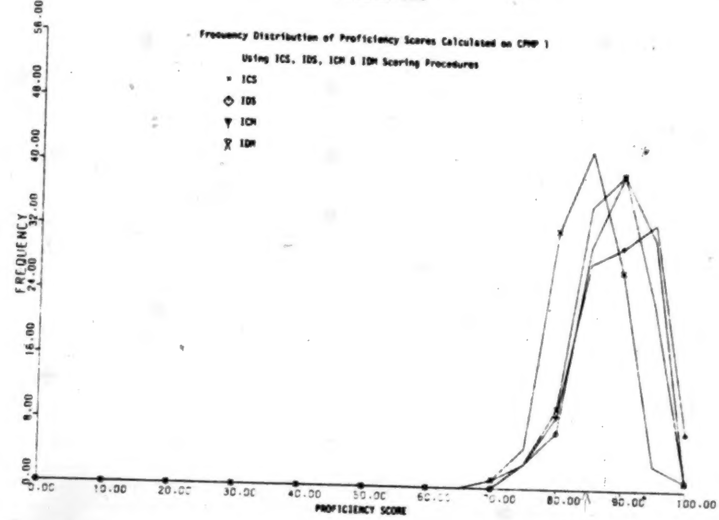
NOTE: Steps 3 and 4 may be carried out simultaneously.

APPENDIX F·1



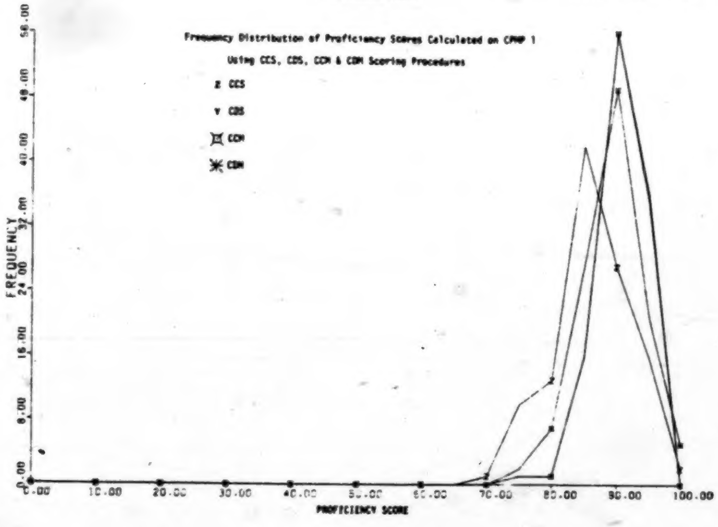
Frequency Distribution of Proficiency Scores Calculated on CPMP 1

Using Author, GCS, GDS, Mc Scoring Procedures



Frequency Distribution of Proficiency Scores Calculated on CPMP 1

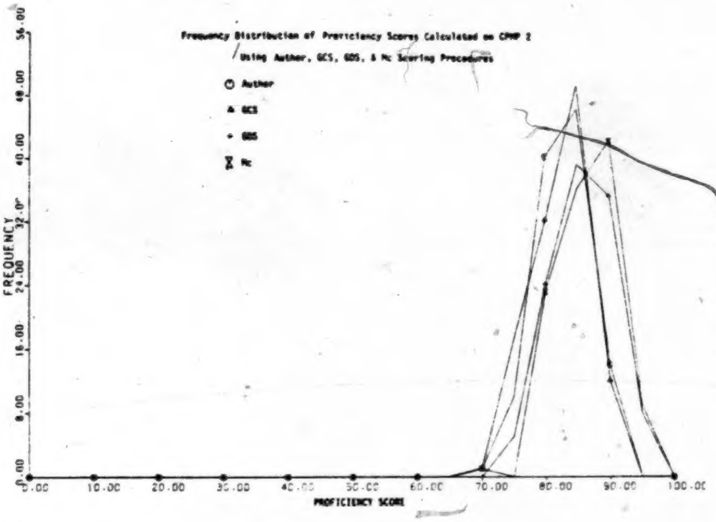
Using ICS, IDS, ICM & IDM Scoring Procedures

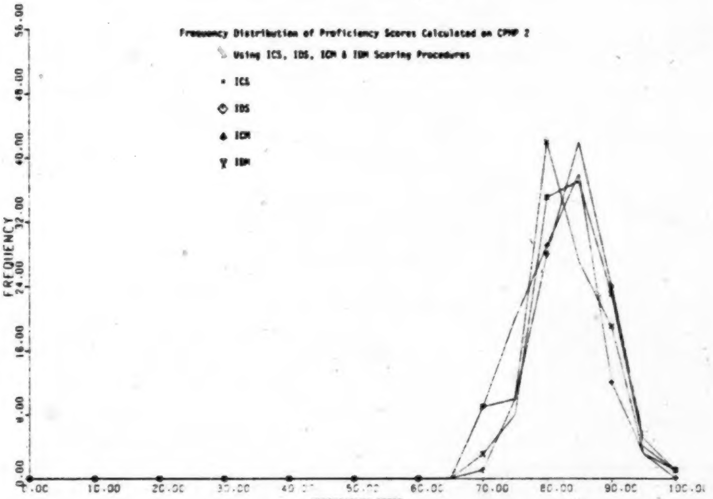


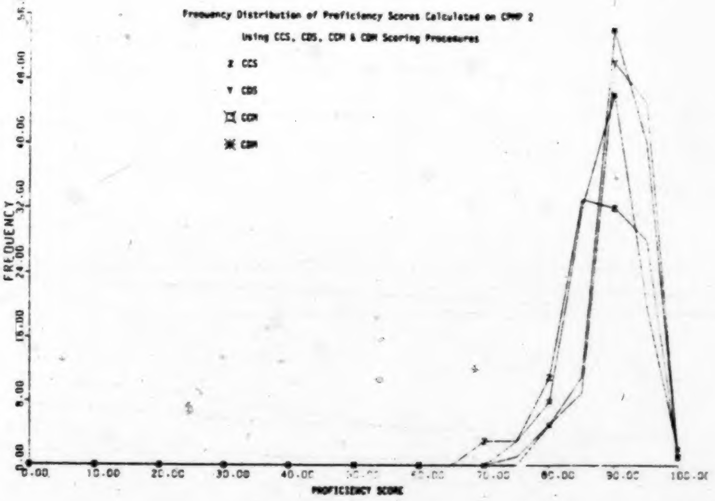
Frequency Distribution of Proficiency Scores Calculated on CPMP 1

Using CCS, CDS, CCM & CDM Scoring Procedures

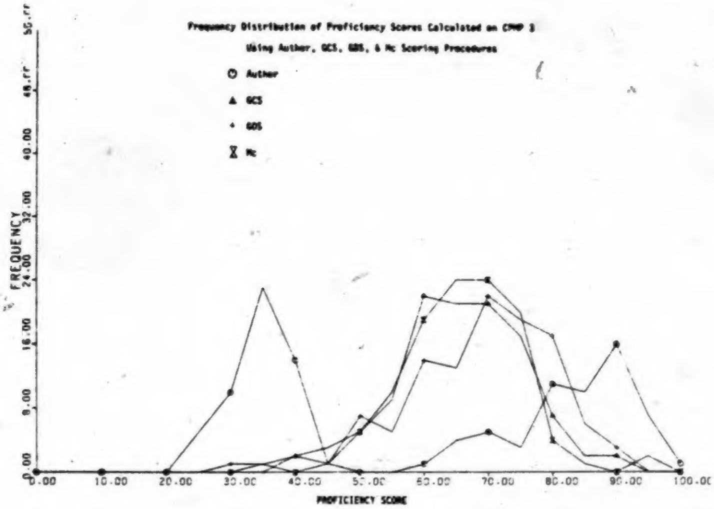
APPENDIX F.2

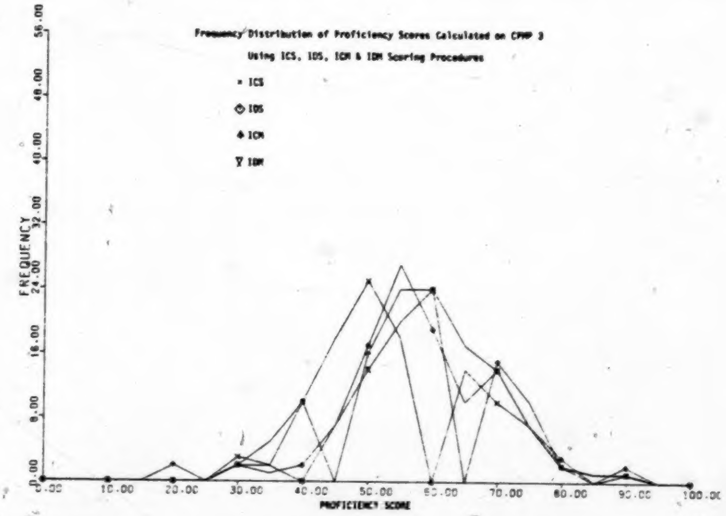

Frequency Distribution of Proficiency Scores Calculated on CPMP 2
Using Author, GCS, GDS, & Mc Scoring Procedures

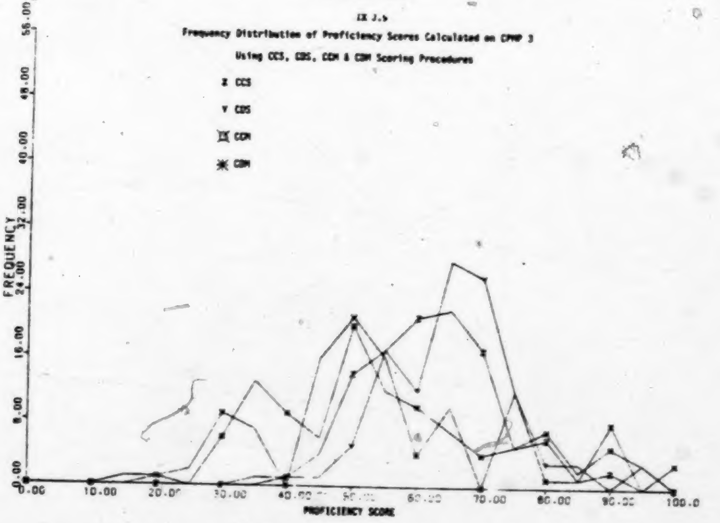

Frequency Distribution of Proficiency Scores Calculated on CPMP 2
Using ICS, IDS, ICM & IDM Scoring Procedures


Frequency Distribution of Proficiency Scores Calculated on CPMP 2
Using CCS, CDS, CCM & CDM Scoring Procedures

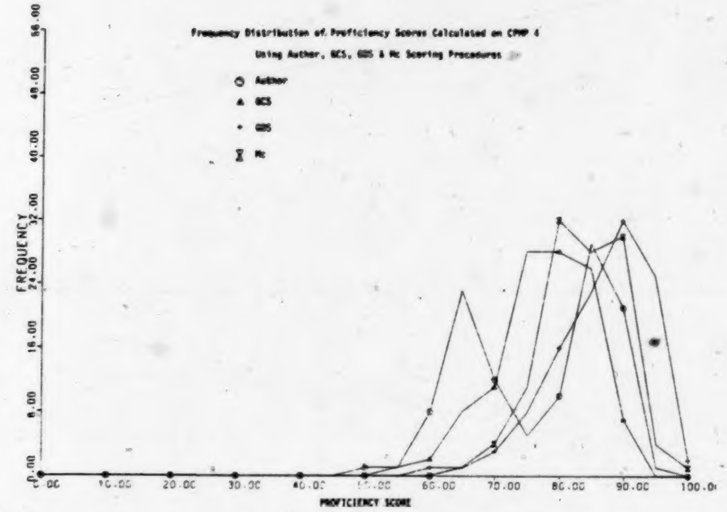
APPENDIX F.3

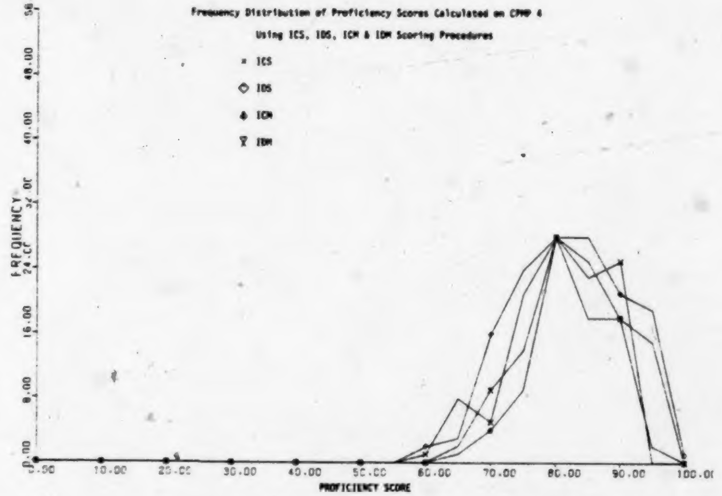

Frequency Distribution of Proficiency Scores Calculated on CPMP 3
Using Auther, GCS, GDS, & Mc Scoring Procedures
○ Auther
▲ GCS
+ GDS
X Mc
FREQUENCY
0.00
8.00
16.00
24.00
32.00
40.00
48.00
56.00
0.00
10.00
20.00
30.00
40.00
50.00
60.00
70.00
80.00
90.00
100.00
PROFICIENCY SCORE

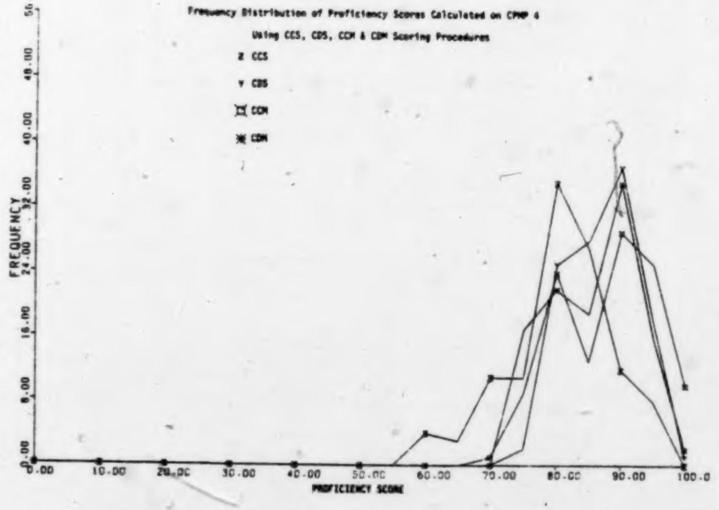

Frequency Distribution of Proficiency Scores Calculated on CPMP 3
Using ICS, IDS, ICM & IDM Scoring Procedures
× ICS
◇ IDS
◆ ICM
▽ IDM
FREQUENCY
0.00
8.00
16.00
24.00
32.00
40.00
48.00
56.00
0.00
10.00
20.00
30.00
40.00
50.00
60.00
70.00
80.00
90.00
100.00
PROFICIENCY SCORE


IX J.5
Frequency Distribution of Proficiency Scores Calculated on CPMP 3
Using CCS, CDS, CCM & CDM Scoring Procedures
X CCS
Y CDS
⌛ CCM
✱ CDM
FREQUENCY
0.00
8.00
16.00
24.00
32.00
40.00
48.00
56.00
0.00
10.00
20.00
30.00
40.00
50.00
60.00
70.00
80.00
90.00
100.0
PROFICIENCY SCORE

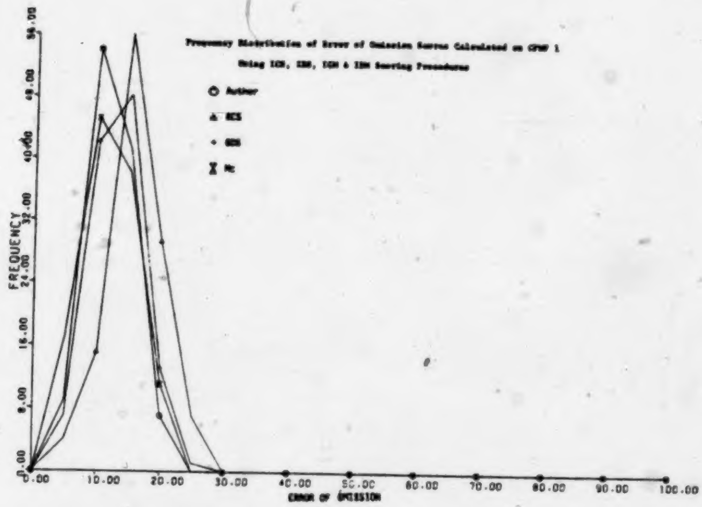
APPENDIX F.4

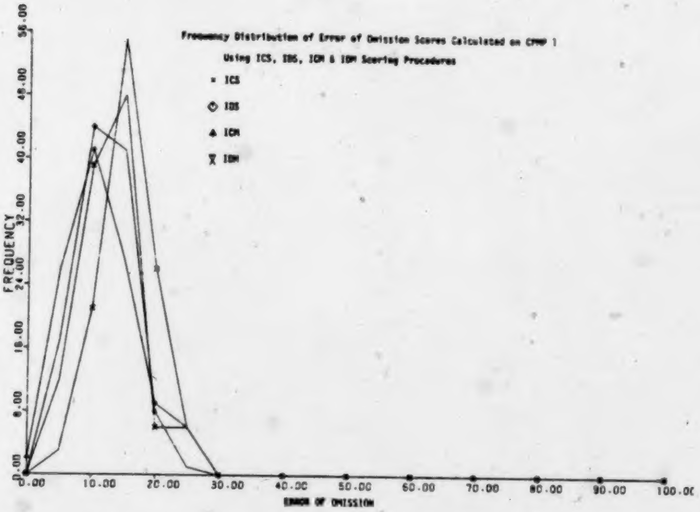

Frequency Distribution of Proficiency Scores Calculated on CPMP 4
Using Author, BCS, GDS & Mc Scoring Procedures
Author
BCS
GDS
Mc
FREQUENCY
0.00
8.00
16.00
24.00
32.00
40.00
48.00
56.00
0.00
10.00
20.00
30.00
40.00
50.00
60.00
70.00
80.00
90.00
100.00
PROFICIENCY SCORE

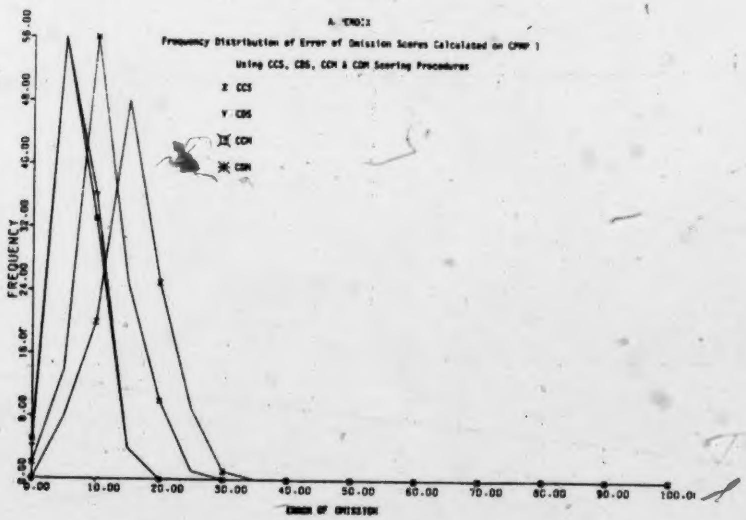

Frequency Distribution of Proficiency Scores Calculated on CPMP 4
Using ICS, IDS, ICM & IDM Scoring Procedures
ICS
IDS
ICM
IDM
FREQUENCY
0.00
8.00
16.00
24.00
32.00
40.00
48.00
56.00
0.00
10.00
20.00
30.00
40.00
50.00
60.00
70.00
80.00
90.00
100.00
PROFICIENCY SCORE


Frequency Distribution of Proficiency Scores Calculated on CPMP 4
Using CCS, CDS, CCM & CDM Scoring Procedures
CCS
CDS
CCM
CDM
FREQUENCY
0.00
8.00
16.00
24.00
32.00
40.00
48.00
56.00
0.00
10.00
20.00
30.00
40.00
50.00
60.00
70.00
80.00
90.00
100.0
PROFICIENCY SCORE

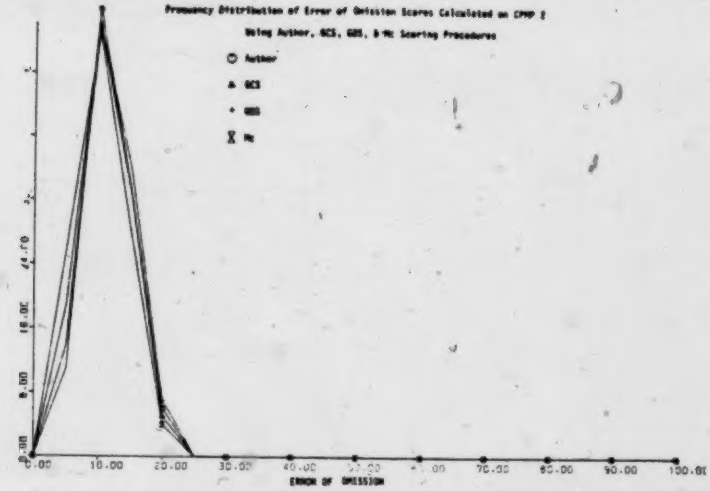
APPENDIX F.5


Frequency Distribution of Error of Omission Scores Calculated on CPMP 1
Using ICS, IDS, ICM & IDM Scoring Procedures
○ Author
▲ RCS
+ BDS
Mc
FREQUENCY
0.00
8.00
16.00
24.00
32.00
40.00
48.00
56.00
0.00
10.00
20.00
30.00
40.00
50.00
60.00
70.00
80.00
90.00
100.00
ERROR OF OMISSION

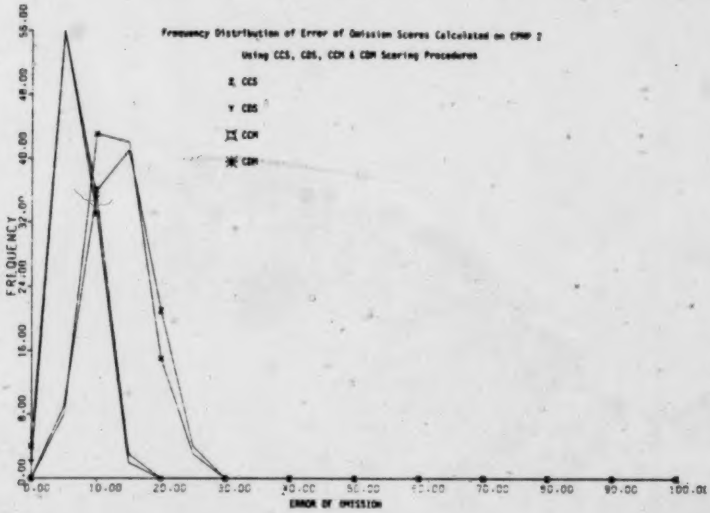

Frequency Distribution of Error of Omission Scores Calculated on CPMP 1
Using ICS, IDS, ICM & IDM Scoring Procedures
× ICS
○ IDS
▲ ICM
IDM
FREQUENCY
0.00
8.00
16.00
24.00
32.00
40.00
48.00
56.00
0.00
10.00
20.00
30.00
40.00
50.00
60.00
70.00
80.00
90.00
100.00
ERROR OF OMISSION


A. 'ENDIX
Frequency Distribution of Error of Omission Scores Calculated on CPMP 1
Using CCS, CDS, CCM & CDM Scoring Procedures
Z CCS
Y CDS
CCM
CDM
FREQUENCY
0.00
8.00
16.00
24.00
32.00
40.00
48.00
56.00
0.00
10.00
20.00
30.00
40.00
50.00
60.00
70.00
80.00
90.00
100.00
ERROR OF OMISSION

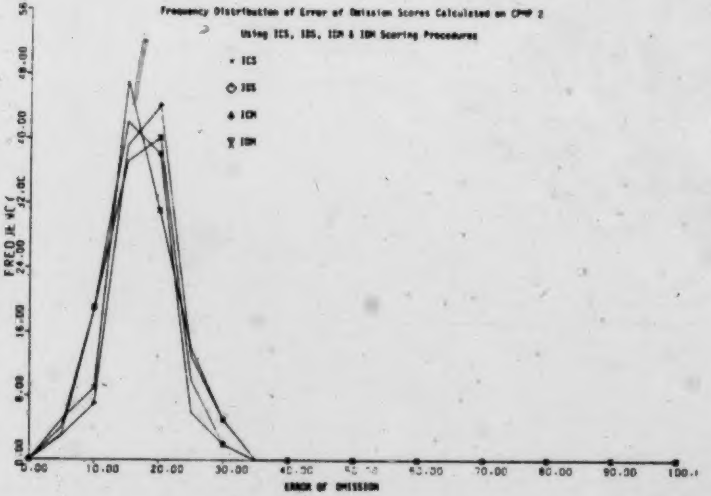
APPENDIX F.6



Frequency Distribution of Error of Omission Scores Calculated on CPMP 2
Using Author, GCS, GDS, & Mc Scoring Procedures

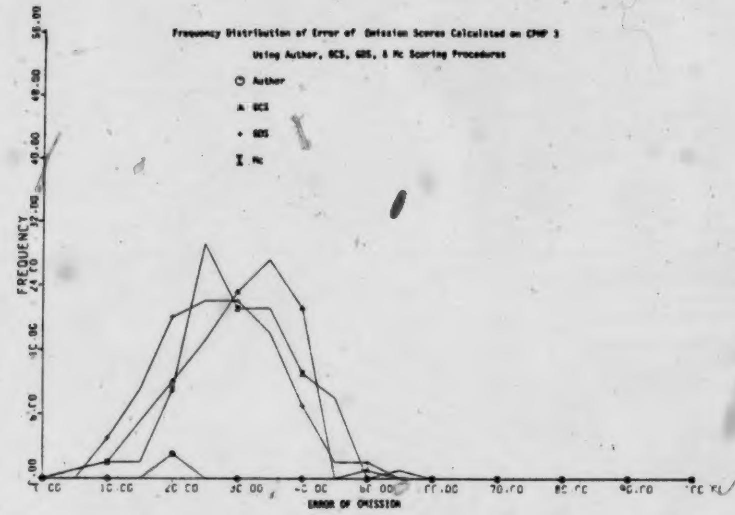


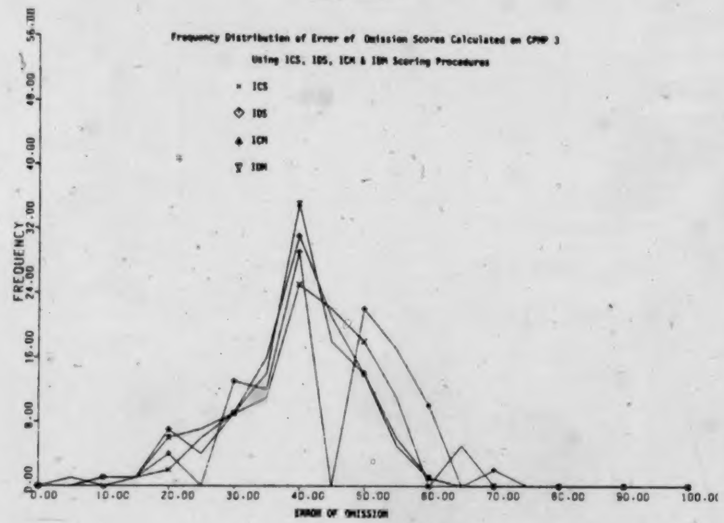
Frequency Distribution of Error of Omission Scores Calculated on CPMP 2
Using CCS, CDS, CCM & CDM Scoring Procedures

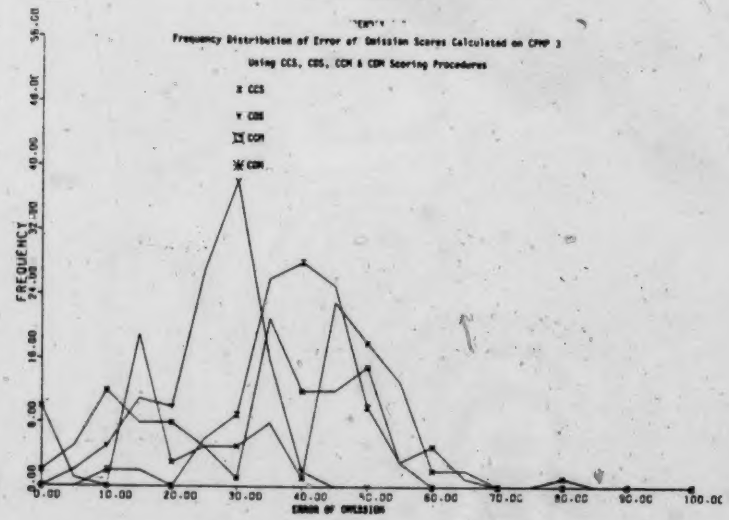


Frequency Distribution of Error of Omission Scores Calculated on CPMP 2
Using ICS, IDS, ICM & IDM Scoring Procedures

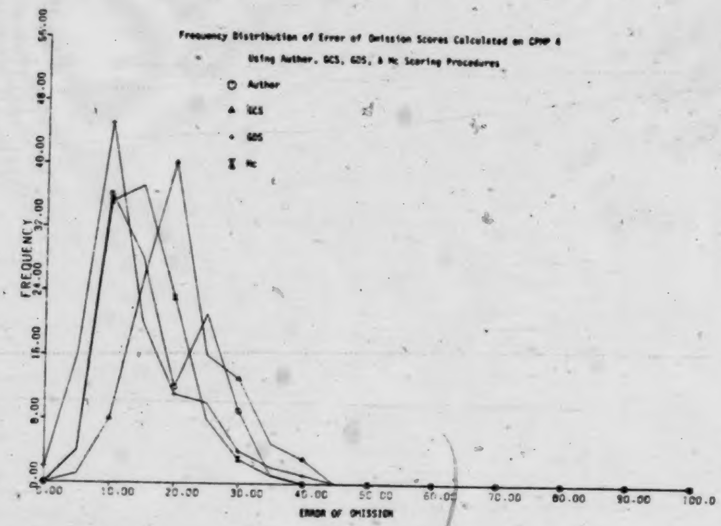
APPENDIX F.7


Frequency Distribution of Error of Omission Scores Calculated on CMP 3
Using Author, BCS, GDS, & Mc Scoring Procedures
○ Author
▲ BCS
◆ GDS
I Mc
FREQUENCY
0.00
8.00
16.00
24.00
32.00
40.00
48.00
56.00
0.00
10.00
20.00
30.00
40.00
50.00
60.00
70.00
80.00
90.00
100.00
ERROR OF OMISSION


Frequency Distribution of Error of Omission Scores Calculated on CMP 3
Using ICS, IDS, ICM & IDM Scoring Procedures
× ICS
◇ IDS
▲ ICM
Y IDM
FREQUENCY
0.00
8.00
16.00
24.00
32.00
40.00
48.00
56.00
0.00
10.00
20.00
30.00
40.00
50.00
60.00
70.00
80.00
90.00
100.00
ERROR OF OMISSION

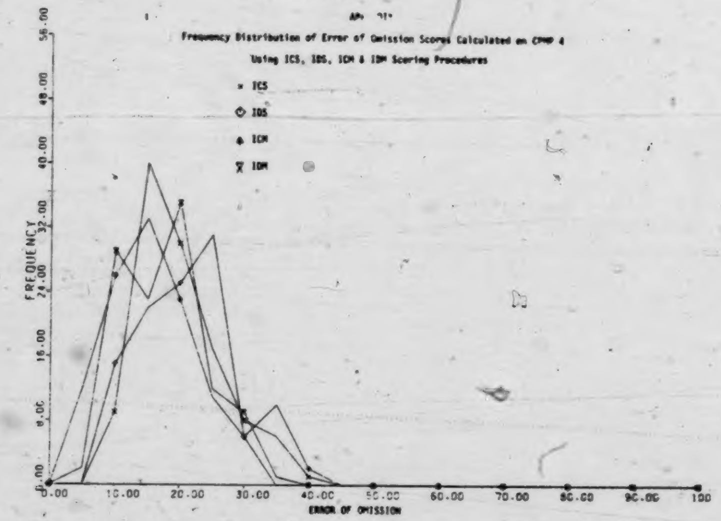

Frequency Distribution of Error of Omission Scores Calculated on CMP 3
Using CCS, CDS, CCM & CDM Scoring Procedures
Z CCS
Y CDS
⌧ CCM
✱ CDM
FREQUENCY
0.00
8.00
16.00
24.00
32.00
40.00
48.00
56.00
0.00
10.00
20.00
30.00
40.00
50.00
60.00
70.00
80.00
90.00
100.00
ERROR OF OMISSION

APPENDIX F.8



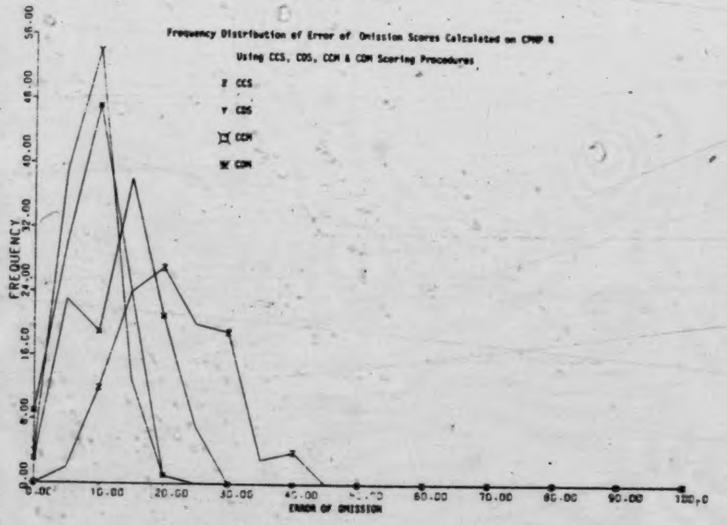
Frequency Distribution of Error of Omission Scores Calculated on CPMP 4

Using Author, GCS, GDS, & Mc Scoring Procedures



Frequency Distribution of Error of Omission Scores Calculated on CPMP 4

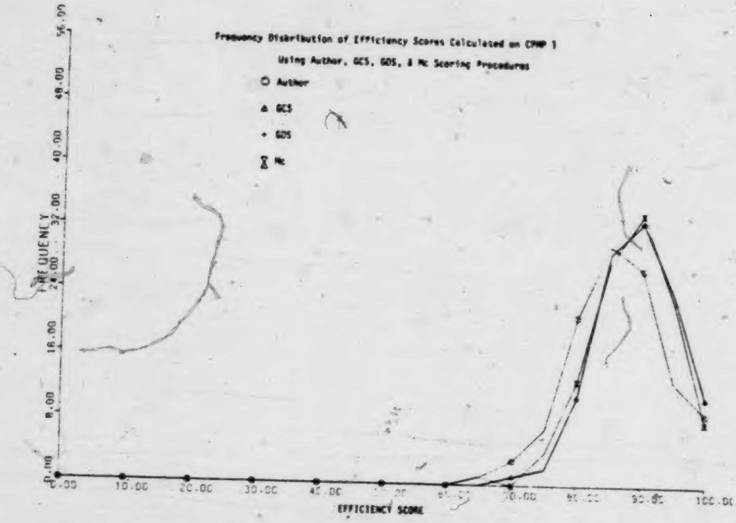
Using ICS, IDS, ICM & IDM Scoring Procedures

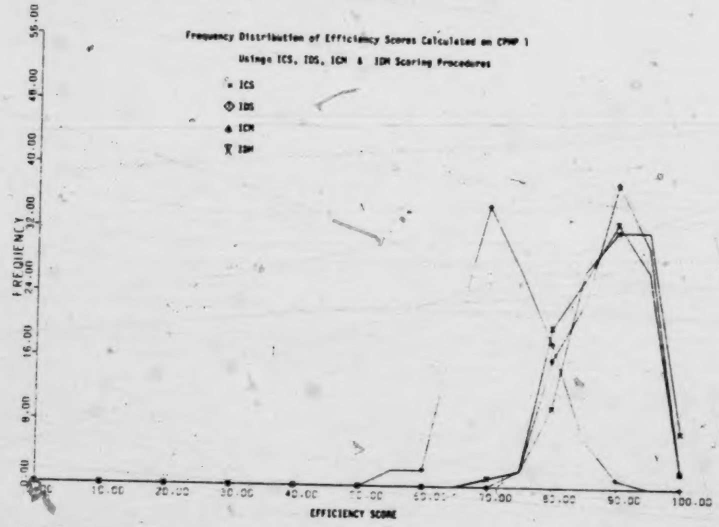


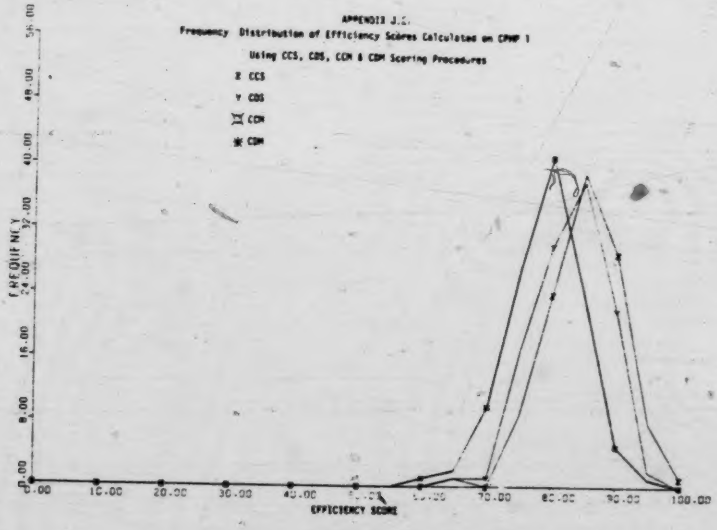
Frequency Distribution of Error of Omission Scores Calculated on CPMP 4

Using CCS, CDS, CCM & CDM Scoring Procedures

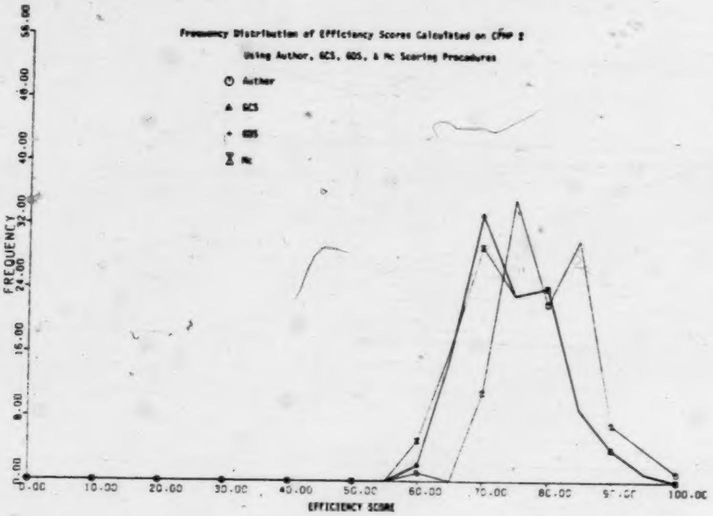
APPENDIX F.9


Frequency Distribution of Efficiency Scores Calculated on CPMP 1
Using Author, GCS, GDS, & Mc Scoring Procedures

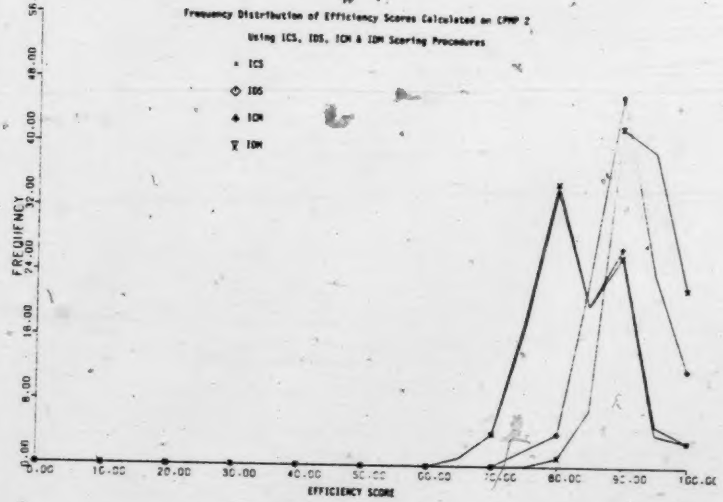

Frequency Distribution of Efficiency Scores Calculated on CPMP 1
Usinge ICS, IDS, ICM & IDM Scoring Procedures


APPENDIX J.2.
Frequency Distribution of Efficiency Scores Calculated on CPMP 1
Using CCS, CDS, CCM & CDM Scoring Procedures

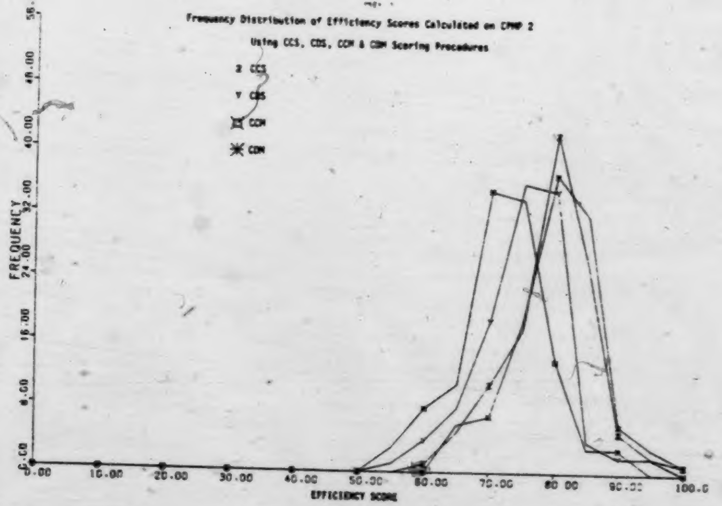
APPENDIX F.10



Frequency Distribution of Efficiency Scores Calculated on CPMP 2
Using Author, GCS, GDS, & Mc Scoring Procedures

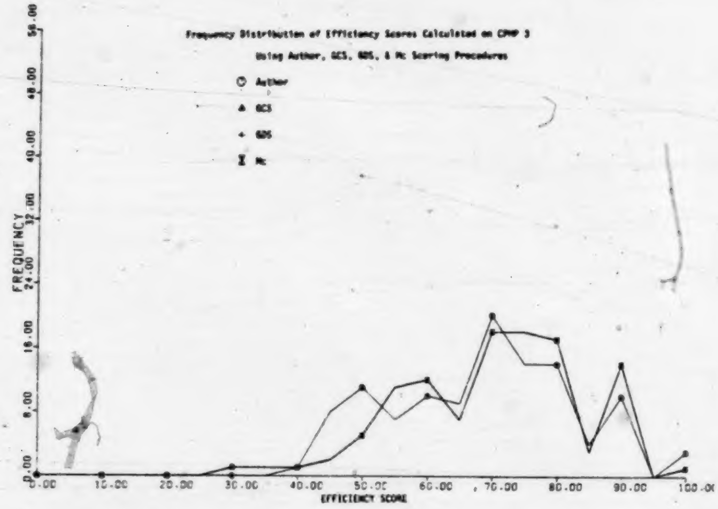


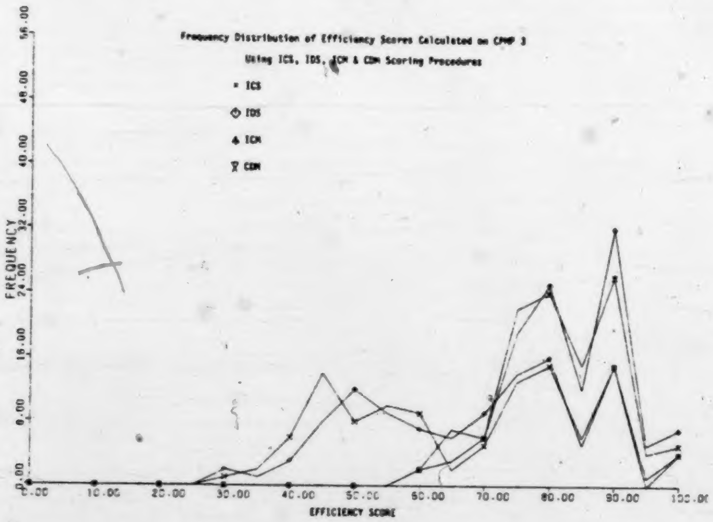
Frequency Distribution of Efficiency Scores Calculated on CPMP 2
Using ICS, IDS, ICM & IDM Scoring Procedures

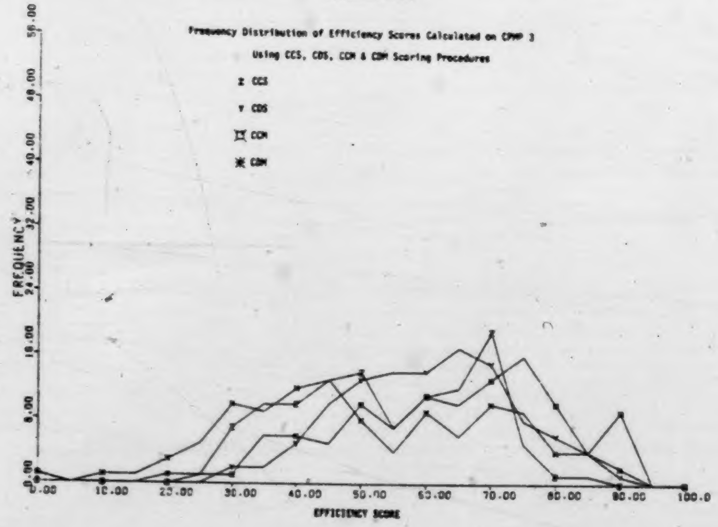


Frequency Distribution of Efficiency Scores Calculated on CPMP 2
Using CCS, CDS, CCM & CDM Scoring Procedures

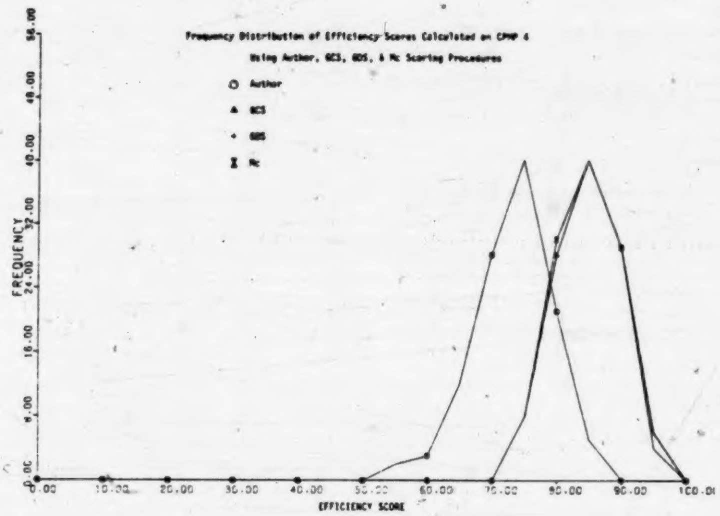
APPENDIX F.11





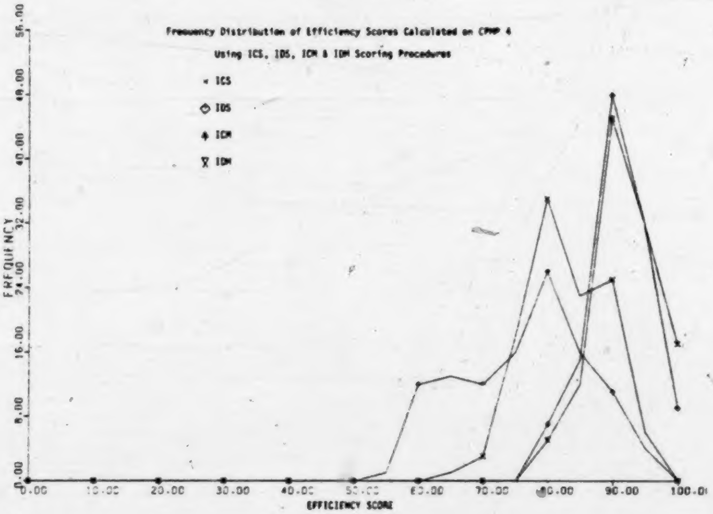


APPENDIX F.12



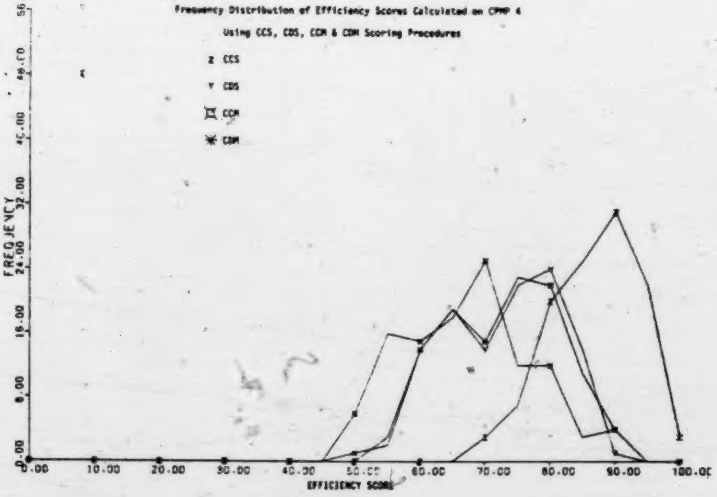
Frequency Distribution of Efficiency Scores Calculated on CPMP 4

Using Author, GCS, GDS, & Mc Scoring Procedures



Frequency Distribution of Efficiency Scores Calculated on CPMP 4

Using ICS, IDS, ICM & IDM Scoring Procedures



Frequency Distribution of Efficiency Scores Calculated on CPMP 4

Using CCS, CDS, CCM & CDM Scoring Procedures

APPENDIX G.1

The X'X matrix of a multivariate analysis with repeated measures is made of four submatrices which are illustrated below:

$$\begin{array}{c} \\ 1 \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ n+m \end{array}
\left[\begin{array}{cccccccccc:cccccc}
1 & & & & & & & & & n & 1 & & & & & m \\
m & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & \cdots & 1 & 1 \\
0 & m & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & \cdots & 1 & 1 \\
0 & 0 & m & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & \cdots & 1 & 1 \\
0 & 0 & 0 & m & \mathbf{a} & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & \mathbf{b} & 1 & 1 \\
\vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\
0 & 0 & 0 & 0 & & m & 0 & 0 & 0 & 0 & 1 & 1 & 1 & & 1 & 1 \\
0 & 0 & 0 & 0 & \cdots & 0 & m & 0 & 0 & 0 & 1 & 1 & 1 & \cdots & 1 & 1 \\
0 & 0 & 0 & 0 & \cdots & 0 & 0 & m & 0 & 0 & 1 & 1 & 1 & \cdots & 1 & 1 \\
0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & m & 0 & 1 & 1 & 1 & \cdots & 1 & 1 \\
0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & m & 1 & 1 & 1 & \cdots & 1 & 1 \\
\hdashline
1 & 1 & 1 & 1 & \cdots & 1 & 1 & 1 & 1 & 1 & n & 0 & 0 & \cdots & 0 & 0 \\
1 & 1 & 1 & 1 & \cdots & 1 & 1 & 1 & 1 & 1 & 0 & n & 0 & \cdots & 0 & 0 \\
1 & 1 & 1 & 1 & \mathbf{b} & 1 & 1 & 1 & 1 & 1 & 0 & 0 & n & \mathbf{c} & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\
1 & 1 & 1 & 1 & & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & & n & 0 \\
1 & 1 & 1 & 1 & \cdots & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & \cdots & 0 & n
\end{array}\right]
\begin{array}{c} \\ 1 \\ \\ \\ \\ \\ \\ \\ \\ \\ n \\ 1 \\ \\ \\ \\ \\ m \end{array}
= X'X$$

Figure 1: Structure of the X'X matrix.

where n = number of repeated cells (i.e., students), and

m = number of repetitions (i.e., scoring procedures).

Submatrix a, a square matrix of dimension n X n, contains zeroes on the off-diagonal and 'm' elements on the diagonals. Submatrix b, a rectangular matrix of dimension m X m, contains ones throughout. Submatrix c, a square matrix of dimension m X m, contains zeros in

APPENDIX G.2

the off-diagonal and n elements in the diagonal positions. Computing the generalized inverse of the X'X matrix gave the following results.

$$
\begin{array}{c}
\\ 1 \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ n+m
\end{array}
\left[\begin{array}{cccccccccc|cccccc}
1 & & & & & & & & & n & 1 & & & & & m \\
a & b & b & b & \cdots & b & b & b & b & b & c & c & c & \cdots & c & c \\
b & a & b & b & \cdots & b & b & b & b & b & c & c & c & \cdots & c & c \\
b & b & a & b & \cdots & b & b & b & b & b & c & c & c & \cdots & c & c \\
b & b & b & a & \mathbf{a'} & b & b & b & b & b & c & c & c & \mathbf{b'} & c & c \\
\vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\
b & b & b & b & \cdots & a & b & b & b & b & c & c & c & \cdots & c & c \\
b & b & b & b & \cdots & b & a & b & b & b & c & c & c & \cdots & c & c \\
b & b & b & b & \cdots & b & b & a & b & b & c & c & c & \cdots & c & c \\
b & b & b & b & \cdots & b & b & b & a & b & c & c & c & \cdots & c & c \\
b & b & b & b & \cdots & b & b & b & b & a & c & c & c & \cdots & c & c \\
\hline
c & c & c & c & \cdots & c & c & c & c & c & d & e & e & \cdots & e & e \\
c & c & c & c & \mathbf{b'} & c & c & c & c & c & e & d & e & \cdots & e & e \\
c & c & c & c & \cdots & c & c & c & c & c & e & e & d & \mathbf{c'} & e & e \\
\vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\
c & c & c & c & \cdots & c & c & c & c & c & e & e & e & \cdots & d & e \\
c & c & c & c & \cdots & c & c & c & c & c & e & e & e & \cdots & e & d
\end{array}\right]
\begin{array}{c}
\\ 1 \\ \\ \\ \\ \\ \\ \\ \\ \\ n \\ 1 \\ \\ \\ \\ \\ m
\end{array}
= (X'X)^{-}
$$

Figure 1: Structure of the generalized inverse of the X'X matrix

Submatrix a', contained the constant, a, in the diagonal and the constant, b, in the off-diagonal. Submatrix b', contained only the constant, c. Submatrix c', contained the constant, d, in the diagonal and the constant, e, in the off-diagonal.

Analyses of the generalized inverse of several matrices yielded the following five equations:

APPENDIX G.3

$$a + (n - 1)b = mc$$

$$d + (m - 1)e = nc$$

$$m(a - b) = 1$$

$$n(d - e) = 1$$

$$c(m + n)^2 = 1$$

Since the five equations contained five unknowns (*i.e.*, a - e), a solution of unknowns was calculated. The validity of the algebraic solution of the generalized inverse of X'X was ensured by the pre and post multiplication of the generalized inverse matrix by the original matrix which yielded the original matrix (*i.e.*, $(X'X)\ (X'X)^{-}\ (X'X) = (X'X)$).

APPENDIX H

One-Way Multivariate Analysis
With Repeated Measures Over Scoring Procedures

A. Sums of Squares and Cross-Products Due to Total (Y'Y matrix)

	CPMP 1			CPMP 2			CPMP 3			CPMP 4		
	Prof*	E of C**	Eff***	Prof	E of C	Eff	Prof	E of C	Eff	Prof	E of C	Eff
Prof	9954298.5	1194599.9	9592619.2	9672329.5	1301383.7	9037326.3	6807983.9	3514179.3	7607223.0	9358655.6	1677022.5	9098319.3
E of C		185688.0	1191063.2	1182504.2	172076.2	1117943.1	835276.4	445973.2	936980.7	1146329.9	225264.5	1143671.4
Eff			9347324.4	9334033.9	1266866.1	8741190.9	6600962.0	3377321.0	7388424.2	9033398.3	1638988.0	8845230.3
Prof				9461463.9	1235721.9	8807019.5	6630121.1	3425041.3	7378964.5	9124085.1	1626322.9	8862284.5
E of C					214259.5	1206430.4	889739.9	487565.4	1025362.1	1221102.1	241708.4	1215819.1
Eff						8325572.7	6184125.1	3323768.6	6975031.3	8504041.5	1552711.8	8823738.2
Prof							4963958.2	2235169.4	5378941.9	6422456.6	1154203.0	6278467.9
E of C								1552619.9	2614395.6	3328037.1	610272.0	3262414.0
Eff									6218668.2	7169463.2	1299110.0	7068250.6
Prof										8908799.0	1514880.7	8608636.3
E of C											366338.6	1559006.7
Eff												8507199.2

* Prof = Proficiency Score
** E of C = Error of Commission Score
*** Eff = Efficiency Score

APPENDIX I

One-Way Multivariate Analysis
With Repeated Measures Over Scoring Procedures

B. Sums of Squares and Cross-Products Due to Model ($\hat{Y}'\hat{Y}$ matrix)

	CPMP 1			CPMP 2			CPMP 3			CPMP 4		
	Prof*	E of C**	Eff***	Prof	E of C	Eff	Prof	E of C	Eff	Prof	E of C	Eff
Prof	9949293.1	1197547.4	9558826.6	9671900.6	1301587.8	9037187.6	6808204.4	3514952.5	7607174.1	9357952.5	1677312.7	9098140.6
E of C		183233.9	1193703.5	118598.5	172026.6	1118008.5	836274.1	444800.5	937566.6	1146886.6	224956.8	1143975.0
Eff			9338255.6	9333540.6	1266953.4	8740586.4	6598944.6	3379054.6	7387712.4	9033194.4	1639152.4	8844913.0
Prof				9451446.1	1241244.9	8798646.9	6629718.0	3423068.0	7379035.1	9123699.6	1626874.1	8862129.1
E of C					210163.7	1212002.2	889367.6	488743.5	1025036.6	1221417.2	241232.4	1215979.7
Eff						8310630.7	6183302.4	3232727.7	6974298.4	8504996.7	1552995.8	8322686.1
Prof							4807986.9	2314341.9	5326644.2	6425990.7	1152792.8	6280374.7
E of C								1459792.6	2665424.7	3323660.4	611949.8	3259951.3
Eff									6093482.3	7171507.5	1298453.9	7067819.5
Prof										8888800.1	1527749.6	8594694.5
E of C											355783.6	1568777.2
Eff												8479163.0

* Prof = Proficiency Score
** E of C = Error of Commission Score
*** Eff = Efficiency Score

APPENDIX J

One-Way Multivariate Analysis
With Repeated Measures Over Scoring Procedures

C. Sums of Squares Due to Model Corrected for Sums of Squares Due to Means

	CPMP 1			CPMP 2			CPMP 3			CPMP 4		
	Prof*	E of C**	Eff***	Prof	E of C	Eff	Prof	E of C	Eff	Prof	E of C	Eff
Prof	31454.5	-27844.4	3456.8	8742.2	-3011.6	8650.1	4547.1	-104.4	7313.7	5709.5	-9105.1	-10971.1
E of C		31831.4	9389.7	-11326.5	10837.5	2493.7	-4347.1	10500.1	-1429.0	-8623.4	16592.6	18505.0
Eff			74209.6	-5686.3	6087.2	14706.3	23361.6	-18169.2	42616.9	-5539.8	9267.3	41159.8
Prof				36428.1	-29853.7	1953.2	771.8	-1725.5	-25668.3	11612.9	-16238.2	-13069.5
E of C					38555.8	24382.1	-5590.2	26370.5	25345.7	-8783.4	19399.8	17760.8
Eff						91653.4	-10292.6	32855.0	55893.3	-8660.4	17793.8	30359.5
Prof							140644.2	-96994.0	113124.6	10333.1	-4093.2	31505.7
E of C								213994.6	-38099.6	9060.9	14253.7	31521.8
Eff									269846.8	5052.9	6182.4	87672.0
Prof										69897.9	-62495.0	5058.3
E of C											69027.1	19874.4
Eff												112832.6

* Prof = Proficiency Score
** E of C = Error of Commission Score
*** Eff = Efficiency Score

APPENDIX K

One-Way Multivariate Analysis
With Repeated Measures Over Scoring Procedures

E. Sums of Squares and Cross-Products Due to Scoring Procedures (H matrix)

	CPMP 1			CPMP 2			CPMP 3			CPMP 4		
	Prof*	E of C**	Eff***	Prof	E of C	Eff	Prof	E of C	Eff	Prof	E of C	Eff
Prof	6572.7	-7778.0	-9743.6	5676.2	-3515.6	-60.4	-3458.7	2516.7	-3342.7	6810.4	-9392.7	-10891.4
E of C		11468.8	9941.9	-8327.9	9343.0	6523.8	1826.8	7949.5	4369.1	-7647.3	14558.3	17128.7
Eff			37187.7	-4546.7	-75.4	3682.2	16394.5	-15747.0	29029.4	-1490.1	3929.9	35741.9
Prof				13947.7	-12407.5	-13898.9	2988.6	-3597.4	-22276.5	11484.1	-15352.8	-13265.0
E of C					19970.2	26176.5	-9946.3	28799.8	16380.9	-6817.4	16639.9	17742.8
Eff						52941.0	-16414.2	33978.1	41645.7	-4257.2	13907.1	30198.8
Prof							28281.4	-22543.9	7191.7	2574.3	-2797.7	14866.0
E of C								150480.5	9752.9	11415.7	15078.8	37979.7
Eff									112377.8	-8217.1	11062.5	62738.0
Prof										16994.2	-16510.8	20924,9
E of C											25043.8	20924.9
Eff												77442.2

* Prof = Proficiency Score

** E of C = Error of Commission Score

*** Eff = Efficiency Score

APPENDIX L

One-Way Multivariate Analysis
With Repeated Measures Over Scoring Procedures

F. Effects Model (Beta weights due to scoring procedures (B matrix)

CPMP	Score	Scoring Procedure											
		AUTHOR	GCS	GDS	ICS	IDS	ICM	IDM	CCS	CDS	CCM	CDM	Mc
1	Prof*	76.9	73.5	77.2	74.8	79.1	79.4	78.0	75.9	80.7	79.7	81.1	78.0
	E of C**	9.3	13.5	10.5	13.1	9.5	9.2	10.2	13.3	4.2	9.1	4.0	9.6
	Eff***	76.7	79.8	79.8	79.8	79.2	62.6	78.4	75.5	73.3	69.4	69.4	79.1
2	Prof	73.2	72.6	76.3	72.4	74.5	71.5	73.9	78.2	81.4	78.5	81.1	76.9
	E of C	8.9	8.4	9.0	14.9	13.8	15.9	14.3	11.6	4.0	10.7	3.6	7.8
	Eff	70.8	65.5	65.5	73.8	81.3	74.3	83.9	70.3	66.0	71.3	62.8	65.2
3	Prof	52.3	58.5	61.2	47.8	51.7	49.2	52.3	54.3	58.6	48.2	48.1	58.9
	E of C	-2.4	24.5	22.5	38.2	35.0	39.1	34.5	32.6	21.6	30.3	28.9	26.4
	Eff	60.9	62.7	62.7	59.7	75.6	60.9	73.8	45.8	52.6	43.3	55.4	62.7
4	Prof	67.5	68.4	77.3	71.3	75.5	69.4	73.5	71.7	77.7	78.1	76.3	74.5
	E of C	14.2	18.1	11.1	17.1	12.8	19.0	14.7	18.0	5.4	10.0	6.2	12.5
	Eff	64.1	75.6	75.6	73.6	81.7	67.1	82.6	77.6	63.4	58.2	63.5	75.3

* Prof = Proficiency Scores
** E of C = Error of Commission Scores
*** Eff = Efficiency Scores

APPENDIX M

One-Way Multivariate Analysis
With Repeated Measures Over Scoring Procedures

D. Sums of Squares and Cross-Products Due to Error (E matrix)

	CPMP 1			CPMP 2			CPMP 3			CPMP 4		
	Prof*	E of C**	Eff***	Prof	E of C	Eff	Prof	E of C	Eff	Prof	E of C	Eff
Prof	5005.4	-2947.5	3792.6	428.9	-204.1	138.7	-220.5	-773.0	48.9	703.1	-290.2	178.7
E of C		2454.1	-2640.3	-94.3	49.8	-65.4	-997.7	1172.8	-585.9	-556.7	307.7	-303.6
Eff			9068.8	493.3	-87.3	604.6	2017.4	-1733.6	711.8	203.9	-164.5	317.3
Prof				10017.9	-5523.0	8372.6	403.1	1973.3	-70.6	385.5	-551.2	155.4
E of C					4095.8	-5571.8	372.3	-1178.1	325.5	-315.1	476.0	-160.6
Eff						14942.0	822.7	1040.9	733.0	44.8	-284.0	1052.1
Prof							155971.3	-79172.5	52297.7	-3534.1	1410.2	-1906.8
E of C								92827.3	-41029.0	4376.6	-1677.8	2462.7
Eff									125185.9	-2044.3	656.2	431.1
Prof										19998.9	-12868.9	13941.7
E of C											10554.9	-1906.8
Eff												28036.2

* Prof = Proficiency Scores

** E of C = Error of Commission Scores

*** Eff = Efficiency Scores